Jan-Michael Frahm



## Overview

- Camera model
- Multi-view geometry
- Camera pose estimation
- Feature tracking \& matching
- Robust pose estimation


## A Homogeneous coordinates



18
Introduction to Computer Vision for Robotics


## Properties of affine transformation

Transformation $T_{\text {affine }}$ combines linear mapping and coordinate shift in homogeneous coordinates

- Linear mapping with $\mathrm{A}_{3 \times 3}$ matrix
- coordinate shift with $t_{3}$ translation vector
$M^{\prime}=T_{\text {aftine }} M=\left[\begin{array}{ccc}A_{3 \times 3} & t_{3} \\ 0 & 0 & 0\end{array} 1.1\right] M \quad T_{\text {affine }}=\left[\begin{array}{cccc}a_{11} & a_{12} & a_{13} & t_{x} \\ a_{21} & a_{22} & a_{23} & t_{y} \\ a_{31} & a_{32} & a_{33} & t_{z} \\ 0 & 0 & 0 & 1\end{array}\right]$
- Parallelism is preserved
- ratios of length, area, and volume are preserved
- Transformations can be concatenated:

$$
\text { if } M_{1}=T_{1} M \text { and } M_{2}=T_{2} M_{1} \Rightarrow M_{2}=T_{2} T_{1} M=T_{21} M
$$

## Projective geometry

- Projective space $P^{2}$ is space of rays emerging from $O$
- view point $O$ forms projection center for all rays
- rays $v$ emerge from viewpoint into scene
- ray $g$ is called projective point, defined as scaled $v: g=/ v$


Introduction to Computer Vision for Robotics

## 1 <br> Projective and homogeneous points

- Given: Plane $\Pi$ in $R^{2}$ embedded in $P^{2}$ at coordinates $w=1$
- viewing ray $g$ intersects plane at $v$ (homogeneous coordinates)
- all points on ray $g$ project onto the same homogeneous point $v$
- projection of $g$ onto $\Pi$ is defined by scaling $v=g / l=g / w$



## Finite and infinite points

- All rays $g$ that are not parallel to $\Pi$ intersect at an affine point $v$ on $\Pi$.
- The ray $g(w=0)$ does not intersect $\Pi$. Hence $v_{\infty}$ is not an affine point but a direction. Directions have the coordinates $(x, y, z, 0)^{\top}$
- Projective space combines affine space with infinite points (directions).


Introduction to Computer Vision for Robotics

## Affine and projective transformations

- Affine transformation leaves infinite points at infinity

$$
M_{\infty}^{\prime}=T_{\text {affine }} M_{\infty} \quad \Rightarrow\left[\begin{array}{c}
X^{\prime} \\
Y^{\prime} \\
Z^{\prime} \\
0
\end{array}\right]=\left[\begin{array}{cccc}
a_{11} & a_{12} & a_{13} & t_{x} \\
a_{21} & a_{22} & a_{23} & t_{y} \\
a_{31} & a_{32} & a_{33} & t_{z} \\
0 & 0 & 0 & 1
\end{array}\right]\left[\begin{array}{l}
X \\
Y \\
Z \\
0
\end{array}\right]
$$

- Projective transformations move infinite points into finite affine space
$M^{\prime}=T_{\text {projective }} M_{\infty} \Rightarrow\left[\begin{array}{c}x_{p} \\ y_{p} \\ Z_{p} \\ 1\end{array}\right]=\lambda\left[\begin{array}{c}X^{\prime} \\ Y^{\prime} \\ Z^{\prime} \\ w^{\prime}\end{array}\right]=\lambda\left[\begin{array}{llll}a_{11} & a_{12} & a_{13} & t_{x} \\ a_{21} & a_{22} & a_{23} & t_{y} \\ a_{31} & a_{32} & a_{33} & t_{z} \\ w_{41} & w_{42} & w_{43} & w_{44}\end{array}\right]\left[\begin{array}{c}X \\ Y \\ Z \\ 0\end{array}\right]$

[^0]We can see this intersection due to perspective projection!

## Pinhole Camera (Camera obscura)



Camera obscura


Interior of camera obscura
(Sunday Magazine, 1838) (France, 1830)

Introduction to Computer Vision for Robotics



## object

## A <br> Perspective projection

- Perspective projection in $\mathrm{P}^{3}$ models pinhole camera:
- scene geometry is affine $P^{3}$ space with coordinates $M=(X, Y, Z, 1)^{T}$
- camera focal point in $O=(0,0,0,1)^{\top}$, camera viewing direction along $Z$
- image plane ( $\mathrm{x}, \mathrm{y}$ ) in $\Pi\left(\mathrm{P}^{2}\right)$ aligned with $(\mathrm{X}, \mathrm{Y})$ at $\mathrm{Z}=Z_{0}$
- Scene point $M$ projects onto point $M_{p}$ on plane surface

ntroduction to Computer Vision for Robotics


## A <br> Projective Transformation

- Projective Transformation maps $M$ onto $M_{p}$ in $\mathrm{P}^{3}$ space


$$
\left.\begin{array}{rl}
\rho M_{p}=T_{p} M \Rightarrow \rho & {\left[\begin{array}{c}
x_{p} \\
y_{p} \\
Z_{0} \\
1
\end{array}\right]}
\end{array}\right]=\left[\begin{array}{c}
X \\
Y \\
Z \\
W
\end{array}\right]=\left[\begin{array}{llll}
1 & 0 & 0 & 0 \\
0 & 1 & 0 & 0 \\
0 & 0 & 1 & 0 \\
0 & 0 & \frac{1}{z_{0}} & 0
\end{array}\right]\left[\begin{array}{c}
X \\
Y \\
Z \\
1
\end{array}\right]
$$

- Projective Transformation linearizes projection

Introduction to Computer Vision for Robotics

## A <br> Perspective Projection

Dimension reduction from $\mathrm{P}^{3}$ into $\mathrm{P}^{2}$ by projection onto $\Pi\left(\mathrm{P}^{2}\right)$


Perspective projection $P_{0}$ from $\mathrm{P}^{3}$ onto $\mathrm{P}^{2}$ :

$$
\rho m_{p}=D_{p} T_{p} M=P_{0} M \Rightarrow \rho\left[\begin{array}{c}
x \\
y \\
1
\end{array}\right]=\left[\begin{array}{llll}
1 & 0 & 0 & 0 \\
0 & 1 & 0 & 0 \\
0 & 0 & \frac{1}{z_{0}} & 0
\end{array}\right]\left[\begin{array}{c}
X \\
Y \\
Z \\
1
\end{array}\right], \rho=\frac{Z}{Z_{0}}
$$

Introduction to Computer Vision for Robotics

## A Image plane and image sensor

- A sensor with picture elements (Pixel) is added onto the image plane

- Pixel coordinates are related to image coordinates by affine transformation $K$ with five parameters:
- Image center cat optical axis
- distance $Z_{p}$ (focal length) and Pixel size determines pixel resolution $f_{x}, f_{y}$

$$
K=\left[\begin{array}{ccc}
f_{x} & s & c_{x} \\
0 & f_{y} & c_{y} \\
0 & 0 & 1
\end{array}\right]
$$

- image skew s to model angle between pixel rows and columns

Introduction to Computer Vision for Robotics


## Projection matrix $P$

- Camera projection matrix P combines:
- inverse affine transformation $T_{\text {cam }}{ }^{-1}$ from general pose to origin
- Perspective projection $P_{0}$ to image plane at $Z_{0}=1$
- affine mapping $K$ from image to sensor coordinates
scene pose transformation: $T_{\text {scene }}=\left[\begin{array}{cc}R^{T} & -R^{T} C \\ 0^{T} & 1\end{array}\right]$
projection: $P_{0}=\left[\begin{array}{llll}1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0\end{array}\right]=\left[\begin{array}{ll}I & 0\end{array}\right] \quad$ sensor calibration: $K=\left[\begin{array}{ccc}f_{x} & s & c_{x} \\ 0 & f_{y} & c_{y} \\ 0 & 0 & 1\end{array}\right]$
$\Rightarrow \rho m=P M, \quad P=K P_{0} T_{\text {scene }}=K\left[\begin{array}{ll}R^{T} & -R^{T} C\end{array}\right]$


## 2-view geometry: F-Matrix

Projection onto two views:


## The Fundamental Matrix F

- The projective points $e_{1}$ and $\left(H_{\infty} m_{0}\right)$ define a plane in camera 1 (epipolar plane $\Pi_{e}$ )
- the epipolar plane intersect the image plane 1 in a line (epipolar line $u_{e}$ )
- the corresponding point $m_{1}$ lies on that line: $m_{1}{ }^{\top} u_{e}=0$
- If the points $\left(e_{1}\right),\left(m_{1}\right),\left(H_{\infty} m_{o}\right)$ are all collinear, then the collinearity theorem applies: $\left(m_{1}^{\top} e_{1} \times H_{\infty} m_{0}\right)=0$.
collinearity of $m_{1}, e_{1}, H_{\infty} m_{0} \Rightarrow m_{1}^{T}(\underbrace{\left[e_{1}\right]_{x}} H_{\infty} m_{0})=0$
$[e]_{x}=\left[\begin{array}{ccc}0 & -e_{z} & e_{y} \\ e_{z} & 0 & -e_{x} \\ -e_{y} & e_{x} & 0\end{array}\right]$


| Fundamental Matrix $F$ | Epipolar constraint |
| :---: | :---: |
| $F=\left[e_{1}\right]_{x} H_{\infty}$ | $m_{1}^{\top} F m_{0}=0$ |

## Estimation of $F$ from correspondences

- Given a set of corresponding points, solve linearily for the 9 elements of $F$ in projective coordinates
- since the epipolar constraint is homogeneous up to scale, only eight elements are independent
- since the operator $[e]_{x}$ and hence $F$ have rank $2, F$ has only 7 independent parameters (all epipolar lines intersect at e)
- each correspondence gives 1 collinearity constraint
=> solve F with minimum of 7 correspondences
for $\mathrm{N}>7$ correspondences minimize distance point-line:
$\sum_{n=0}^{N}\left(m_{1, n}^{\top} F m_{0, n}\right)^{2} \Rightarrow \min !$
$m_{1 i}^{\top} F m_{0 i}=0 \quad \operatorname{det}(F)=0$ (Rank 2 constraint)

Introduction to Computer Vision for Robotics


## The Essential Matrix $E$

- $F$ is the most general constraint on an image pair. If the camera calibration matrix K is known, then more constraints are available
- Essential Matrix E

$$
\begin{aligned}
& m_{1}^{T} F m_{0}=\left(K \tilde{m}_{1}\right)^{T} F\left(K \tilde{m}_{0}\right)=\tilde{m}_{1}^{T} \underbrace{\left(K^{T} F K\right)}_{E} \tilde{m}_{0} \\
& E=[e]_{x} R \text { with }[e]_{x}=\left[\begin{array}{ccc}
0 & -e_{z} & e_{y} \\
e_{z} & 0 & -e_{x} \\
-e_{y} & e_{x} & 0
\end{array}\right]
\end{aligned}
$$

- E holds the relative orientation of a calibrated camera pair. It has 5 degrees of freedom: 3 from rotation matrix $\mathrm{R}_{\mathrm{ik}}$, 2 from direction of translation e, the epipole.


## Estimation of $P$ from $E$

- From E we can obtain a camera projection matrix pair: $\mathrm{E}=\mathrm{Udiag}(0,0,1) \mathrm{V}^{\top}$
- $P_{0}=\left[l_{3 \times 3} \mid 0_{3 \times 1}\right]$ and there are four choices for $P_{1}$ :



## 3D Feature Reconstruction

- corresponding point pair $\left(m_{0}, m_{1}\right)$ is projected from 3D feature point M
- $M$ is reconstructed from by $\left(m_{0}, m_{1}\right)$ triangulation
- M has minimum distance of intersection



## 1 <br> Multi View Tracking

- 2D match: Image correspondence ( $m_{1}, m_{i}$ )
- 3D match: Correspondence transfer ( $m i, M$ ) via $P_{1}$
- 3D Pose estimation of $P_{i}$ with $m_{i}-P_{i} M=>$ min.


Minimize lobal reprojection error: $\sum_{i=0}^{N} \sum_{k=0}^{K}\left\|m_{k, i}-P_{i} M_{k}\right\|^{2} \Rightarrow \min !$

## Correspondences matching vs. tracking

- Image-to-image correspondences are essential to 3D reconstruction


Extract features independently and then match by comparing descriptors [Lowe 2004]


Extract features in first images and find same feature back in next view [Lucas \& Kanade 1981] , [Shi \& Tomasi 1994]

- Small difference between frames
- potential large difference overall


## SIFT-detector


. Scale and image-plane-rotation invariant feature descriptor [Lowe 2004]

Introduction to Computer Vision for Robotics


## SIFT-detector

- Image content is transformed into local feature coordinates that are invariant to translation, rotation, scale, and other imaging parameters




## Key point localization

- Detect maxima and minima of difference-of-Gaussian in scale space
- Fit a quadratic to surrounding values for sub-pixel and sub-scale interpolation (Brown \& Lowe, 2002)
- Tavlor expansion around point:

$$
D(\mathbf{x})=D+\frac{\partial D^{T}}{\partial \mathbf{x}} \mathbf{x}+\frac{1}{2} \mathbf{x}^{\mathbf{T}} \frac{\partial^{2} D}{\partial \mathbf{x}^{2}} \mathbf{x}
$$

- Offset of extremum (use finite
 differences for derivatives):

$$
\hat{\mathbf{x}}=-\frac{\partial^{2} D^{-1}}{\partial \mathbf{x}^{2}} \frac{\partial D}{\partial \mathbf{x}}
$$

## Orientation normalization

- Histogram of local gradient directions computed at selected scale
- Assign principal orientation at peak of smoothed histogram

- Each key specifies stable 2D coordinates (x, y, scale, orientation)


Introduction to Computer Vision for Robotics

## Example of keypoint detection

Threshold on value at DOG peak and on ratio of principle curvatures (Harris approach)


## SIFT vector formation

- Thresholded image gradients are sampled over $16 \times 16$ array of locations in scale space
- Create array of orientation histograms
- 8 orientations $\times 4 \times 4$ histogram array $=128$ dimensions
example $2 \times 2$ histogram array


Image gradients


- Estimation of plane from point data




## References

[Lowe 2004] David Lowe, Distinctive Image Features from ScaleInvariant Keypoints, IJCV, 60(2), 2004, pp91-110
[Lucas \& Kanade 1981] Bruce D. Lucas and Takeo Kanade. An Iterative Image Registration Technique with an Application to Stereo Vision. In Proceedings International Joint Conference on Artificial Intelligence, 1981.
[Shi \& Tomasi 1994] Jianbo Shi and Carlo Tomasi, Good Features to Track, IEEE Conference on Computer Vision and Pattern Recognition 1994
[Baker \& Matthews 2004] S. Baker and I. Matthews, Lucas-Kanade 20 Years On: A Unifying Framework.International Journal of Computer Vision, 56(3):221-255, March 2004.
[Fischler \& Bolles] M.A. Fischler and R.C. Bolles, Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography.CACM, 24(6), June'81.

Introduction to Computer Vision for Robotics


## References

- [Mikolajczyk 2003], K.Mikolajczyk, C.Schmid. "A Performance Evaluation of Local Descriptors". CVPR 2003
- [Lindeberg 1998], T. Lindeberg, "Feature detection with automatic scale selection," International Journal of Computer Vision, vol. 30, no. 2, 1998
- [Hartley 2003] R. hartley and A. Zisserman, "Multiple View Geometry in Computer Vision", $2^{\text {nd }}$ edition, Cambridge Press, 2003


[^0]:    Example: Parallel lines intersect at the horizon (line of infinite points).

