

RWTH AACHEN
UNIVERSITY

Computer Vision - Lecture 21

Structure-from-Motion

29.01.2015

Computer Vision WS 14/15

Bastian Leibe
RWTH Aachen
<http://www.vision.rwth-aachen.de>
leibe@vision.rwth-aachen.de

Many slides adapted from Svetlana Lazebnik, Martial Hebert, Steve Seitz

RWTH AACHEN
UNIVERSITY

Announcements

- Exam
 - 1st Date: Monday, 23.02., 13:30 - 17:30h
 - 2nd Date: Thursday, 26.03., 09:30 - 12:30h
 - Closed-book exam, the core exam time will be 2h.
 - Admission requirement: 50% of the exercise points or passed test exam
 - We will send around an announcement with the exact starting times and places by email.
- Test exam
 - Date: Thursday, 05.02., 09:15 - 10:45h, room UMIC 025
 - Core exam time will be 1h
 - Purpose: Prepare you for the questions you can expect.
 - Possibility to collect bonus exercise points!

B. Leibe 2

RWTH AACHEN
UNIVERSITY

Announcements (2)

- Last lecture next Monday: Repetition
 - Summary of all topics in the lecture
 - “Big picture” and current research directions
 - Opportunity to ask questions
- Please use this opportunity and prepare questions!

Computer Vision WS 14/15

B. Leibe 3

RWTH AACHEN
UNIVERSITY

Course Outline

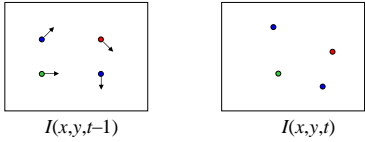
- Image Processing Basics
- Segmentation & Grouping
- Object Recognition
- Local Features & Matching
- Object Categorization
- 3D Reconstruction
 - Epipolar Geometry and Stereo Basics
 - Camera calibration & Uncalibrated Reconstruction
 - Active Stereo
- Motion
 - Motion and Optical Flow
- 3D Reconstruction (Reprise)
 - Structure-from-Motion

Computer Vision WS 14/15

4

RWTH AACHEN
UNIVERSITY

Recap: Estimating Optical Flow



- Given two subsequent frames, estimate the apparent motion field $u(x,y)$ and $v(x,y)$ between them.
- Key assumptions
 - Brightness constancy: projection of the same point looks the same in every frame.
 - Small motion: points do not move very far.
 - Spatial coherence: points move like their neighbors.

Computer Vision WS 14/15

B. Leibe 5

Slide credit: Svetlana Lazebnik

RWTH AACHEN
UNIVERSITY

Recap: Lucas-Kanade Optical Flow

see Exercise 6.4!

- Use all pixels in a $K \times K$ window to get more equations.
- Least squares problem:

$$\begin{bmatrix} I_x(p_1) & I_y(p_1) \\ I_x(p_2) & I_y(p_2) \\ \vdots & \vdots \\ I_x(p_{25}) & I_y(p_{25}) \end{bmatrix} \begin{bmatrix} u \\ v \end{bmatrix} = - \begin{bmatrix} I_t(p_1) \\ I_t(p_2) \\ \vdots \\ I_t(p_{25}) \end{bmatrix} \quad \begin{matrix} A & d = b \\ 25 \times 2 & 2 \times 1 & 25 \times 1 \end{matrix}$$
- Minimum least squares solution given by solution of

$$\begin{bmatrix} \sum I_x I_x & \sum I_x I_y \\ \sum I_x I_y & \sum I_y I_y \end{bmatrix} \begin{bmatrix} u \\ v \end{bmatrix} = - \begin{bmatrix} \sum I_x I_t \\ \sum I_y I_t \end{bmatrix} \quad \begin{matrix} A^T A & & A^T b \\ 2 \times 2 & & 2 \times 1 \end{matrix}$$

Recall the Harris detector!

Computer Vision WS 14/15

B. Leibe 6

Slide adapted from Svetlana Lazebnik

Computer Vision WS 14/15

Recap: Iterative Refinement

- Estimate velocity at each pixel using one iteration of LK estimation.
- Warp one image toward the other using the estimated flow field.
- Refine estimate by repeating the process.
- Iterative procedure
 - Results in subpixel accurate localization.
 - Converges for small displacements.

Slide adapted from Steve Seitz. B. Leibe

Computer Vision WS 14/15

Recap: Coarse-to-fine Estimation

Slide credit: Steve Seitz. B. Leibe

Computer Vision WS 14/15

Recap: Coarse-to-fine Estimation

Slide credit: Steve Seitz. B. Leibe

Computer Vision WS 14/15

Topics of This Lecture

- Structure from Motion (SfM)
 - Motivation
 - Ambiguity
- Affine SfM
 - Affine cameras
 - Affine factorization
 - Euclidean upgrade
 - Dealing with missing data
- Projective SfM
 - Two-camera case
 - Projective factorization
 - Bundle adjustment
 - Practical considerations
- Applications

B. Leibe

Computer Vision WS 14/15

Structure from Motion

- Given: m images of n fixed 3D points

$$x_{ij} = P_i X_j, \quad i = 1, \dots, m, \quad j = 1, \dots, n$$
- Problem: estimate m projection matrices P_i and n 3D points X_j from the mn correspondences x_{ij}

Slide credit: Svetlana Lazebnik. B. Leibe

Computer Vision WS 14/15

What Can We Use This For?

- E.g. movie special effects

[Video](#)

B. Leibe. Video Credit: Stefan Hafeneeger

RWTH AACHEN UNIVERSITY

Structure from Motion Ambiguity

- If we scale the entire scene by some factor k and, at the same time, scale the camera matrices by the factor of $1/k$, the projections of the scene points in the image remain exactly the same:

$$\mathbf{x} = \mathbf{P}\mathbf{X} = \left(\frac{1}{k}\mathbf{P}\right)(k\mathbf{X})$$

⇒ It is impossible to recover the absolute scale of the scene!

Computer Vision WS 14/15 13

Slide credit: Svetlana Lazebnik B. Leibe

RWTH AACHEN UNIVERSITY

Structure from Motion Ambiguity

- If we scale the entire scene by some factor k and, at the same time, scale the camera matrices by the factor of $1/k$, the projections of the scene points in the image remain exactly the same.
- More generally: if we transform the scene using a transformation Q and apply the inverse transformation to the camera matrices, then the images do not change

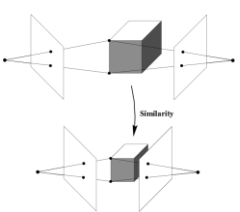
$$\mathbf{x} = \mathbf{P}\mathbf{X} = (\mathbf{P}\mathbf{Q}^{-1})\mathbf{Q}\mathbf{X}$$

Computer Vision WS 14/15 14

Slide credit: Svetlana Lazebnik B. Leibe

RWTH AACHEN UNIVERSITY

Reconstruction Ambiguity: Similarity



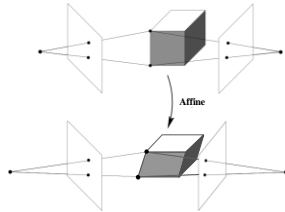
$$\mathbf{x} = \mathbf{P}\mathbf{X} = (\mathbf{P}\mathbf{Q}_S^{-1})\mathbf{Q}_S\mathbf{X}$$

Computer Vision WS 14/15 15

Slide credit: Svetlana Lazebnik B. Leibe Images from Hartley & Zisserman

RWTH AACHEN UNIVERSITY

Reconstruction Ambiguity: Affine



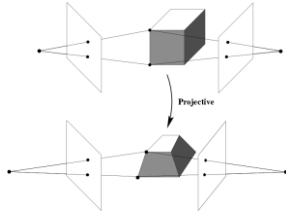
$$\mathbf{x} = \mathbf{P}\mathbf{X} = (\mathbf{P}\mathbf{Q}_A^{-1})\mathbf{Q}_A\mathbf{X}$$

Computer Vision WS 14/15 16

Slide credit: Svetlana Lazebnik B. Leibe Images from Hartley & Zisserman

RWTH AACHEN UNIVERSITY

Reconstruction Ambiguity: Projective



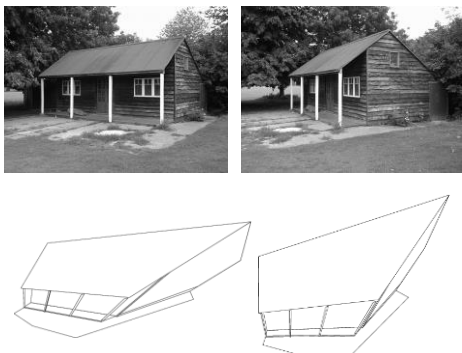
$$\mathbf{x} = \mathbf{P}\mathbf{X} = (\mathbf{P}\mathbf{Q}_P^{-1})\mathbf{Q}_P\mathbf{X}$$

Computer Vision WS 14/15 17

Slide credit: Svetlana Lazebnik B. Leibe Images from Hartley & Zisserman

RWTH AACHEN UNIVERSITY

Projective Ambiguity



Computer Vision WS 14/15 18

Slide credit: Svetlana Lazebnik B. Leibe Images from Hartley & Zisserman

RWTH AACHEN UNIVERSITY

From Projective to Affine

Computer Vision WS 14/15

Slide credit: Svetlana Lazebnik B. Leibe Images from Hartley & Zisserman

19

RWTH AACHEN UNIVERSITY

From Affine to Similarity

Computer Vision WS 14/15

Slide credit: Svetlana Lazebnik B. Leibe Images from Hartley & Zisserman

20

RWTH AACHEN UNIVERSITY

Hierarchy of 3D Transformations

Projective 15dof	$\begin{bmatrix} A & t \\ v^T & v \end{bmatrix}$		Preserves intersection and tangency
Affine 12dof	$\begin{bmatrix} A & t \\ 0^T & 1 \end{bmatrix}$		Preserves parallelism, volume ratios
Similarity 7dof	$\begin{bmatrix} sR & t \\ 0^T & 1 \end{bmatrix}$		Preserves angles, ratios of length
Euclidean 6dof	$\begin{bmatrix} R & t \\ 0^T & 1 \end{bmatrix}$		Preserves angles, lengths

- With no constraints on the camera calibration matrix or on the scene, we get a *projective reconstruction*.
- Need additional information to *upgrade* the reconstruction to affine, similarity, or Euclidean.

Computer Vision WS 14/15 Slide credit: Svetlana Lazebnik B. Leibe

21

RWTH AACHEN UNIVERSITY

Topics of This Lecture

- Structure from Motion (SfM)
 - Motivation
 - Ambiguity
- Affine SfM
 - Affine cameras
 - Affine factorization
 - Euclidean upgrade
 - Dealing with missing data
- Projective SfM
 - Two-camera case
 - Projective factorization
 - Bundle adjustment
 - Practical considerations
- Applications

Computer Vision WS 14/15 B. Leibe

22

RWTH AACHEN UNIVERSITY

Structure from Motion

- Let's start with *affine cameras* (the math is easier)

Computer Vision WS 14/15

Slide credit: Svetlana Lazebnik B. Leibe Images from Hartley & Zisserman

23

RWTH AACHEN UNIVERSITY

Orthographic Projection

- Special case of perspective projection
 - Distance from center of projection to image plane is infinite

- Projection matrix:

$$\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \\ 1 \end{bmatrix} = \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} \Rightarrow (x, y)$$

Computer Vision WS 14/15 Slide credit: Steve Seitz B. Leibe

24

RWTH AACHEN UNIVERSITY

Affine Cameras

Orthographic Projection

Parallel Projection

25

Computer Vision WS 14/15

Slide credit: Svetlana Lazebnik B. Leibe

RWTH AACHEN UNIVERSITY

Affine Cameras

- A general affine camera combines the effects of an affine transformation of the 3D space, orthographic projection, and an affine transformation of the image:

$$P = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} [4 \times 4 \text{ affine}] = \begin{bmatrix} a_{11} & a_{12} & a_{13} & b_1 \\ a_{21} & a_{22} & a_{23} & b_2 \\ 0 & 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} A & b \\ 0 & 1 \end{bmatrix}$$

- Affine projection is a linear mapping + translation in inhomogeneous coordinates

$$\mathbf{x} = \begin{pmatrix} x \\ y \end{pmatrix} = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \end{bmatrix} \begin{pmatrix} X \\ Y \\ Z \end{pmatrix} + \begin{pmatrix} b_1 \\ b_2 \end{pmatrix} = A\mathbf{X} + \mathbf{b}$$

Projection of world origin

26

Computer Vision WS 14/15

Slide credit: Svetlana Lazebnik B. Leibe

RWTH AACHEN UNIVERSITY

Affine Structure from Motion

- Given: m images of n fixed 3D points:
 - $\mathbf{x}_{ij} = A_i \mathbf{X}_j + \mathbf{b}_i, \quad i = 1, \dots, m, j = 1, \dots, n$
- Problem: use the mn correspondences \mathbf{x}_{ij} to estimate m projection matrices A_i and translation vectors \mathbf{b}_i , and n points \mathbf{X}_j
- The reconstruction is defined up to an arbitrary affine transformation Q (12 degrees of freedom):

$$\begin{bmatrix} A & b \\ 0 & 1 \end{bmatrix} \rightarrow \begin{bmatrix} A & b \\ 0 & 1 \end{bmatrix} Q^{-1}, \quad \begin{pmatrix} X \\ 1 \end{pmatrix} \rightarrow Q \begin{pmatrix} X \\ 1 \end{pmatrix}$$
- We have $2mn$ knowns and $8m + 3n$ unknowns (minus 12 dof for affine ambiguity).
 - Thus, we must have $2mn \geq 8m + 3n - 12$.
 - For two views, we need four point correspondences.

27

Computer Vision WS 14/15

Slide credit: Svetlana Lazebnik B. Leibe

RWTH AACHEN UNIVERSITY

Affine Structure from Motion

- Centering: subtract the centroid of the image points

$$\hat{\mathbf{x}}_{ij} = \mathbf{x}_{ij} - \frac{1}{n} \sum_{k=1}^n \mathbf{x}_{ik} = A_i \mathbf{X}_j + \mathbf{b}_i - \frac{1}{n} \sum_{k=1}^n (A_i \mathbf{X}_k + \mathbf{b}_i)$$

$$= A_i \left(\mathbf{X}_j - \frac{1}{n} \sum_{k=1}^n \mathbf{X}_k \right) = A_i \hat{\mathbf{X}}_j$$
- For simplicity, assume that the origin of the world coordinate system is at the centroid of the 3D points.
- After centering, each normalized point $\hat{\mathbf{x}}_{ij}$ is related to the 3D point \mathbf{X}_j by

$$\hat{\mathbf{x}}_{ij} = A_i \mathbf{X}_j$$

28

Computer Vision WS 14/15

Slide credit: Svetlana Lazebnik B. Leibe

RWTH AACHEN UNIVERSITY

Affine Structure from Motion

- Let's create a $2m \times n$ data (measurement) matrix:

$$D = \begin{bmatrix} \hat{\mathbf{x}}_{11} & \hat{\mathbf{x}}_{12} & \dots & \hat{\mathbf{x}}_{1n} \\ \hat{\mathbf{x}}_{21} & \hat{\mathbf{x}}_{22} & \dots & \hat{\mathbf{x}}_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ \hat{\mathbf{x}}_{m1} & \hat{\mathbf{x}}_{m2} & \dots & \hat{\mathbf{x}}_{mn} \end{bmatrix}$$

Points (n)

Cameras ($2m$)

29

Computer Vision WS 14/15

C. Tomasi and T. Kanade, Shape and motion from image streams under orthography: A factorization method, IJCV, 9(2):137-154, November 1992. Slide credit: Svetlana Lazebnik B. Leibe

RWTH AACHEN UNIVERSITY

Affine Structure from Motion

- Let's create a $2m \times n$ data (measurement) matrix:

$$D = \begin{bmatrix} \hat{\mathbf{x}}_{11} & \hat{\mathbf{x}}_{12} & \dots & \hat{\mathbf{x}}_{1n} \\ \hat{\mathbf{x}}_{21} & \hat{\mathbf{x}}_{22} & \dots & \hat{\mathbf{x}}_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ \hat{\mathbf{x}}_{m1} & \hat{\mathbf{x}}_{m2} & \dots & \hat{\mathbf{x}}_{mn} \end{bmatrix} = \begin{bmatrix} A_1 \\ A_2 \\ \vdots \\ A_m \end{bmatrix} \begin{bmatrix} \mathbf{X}_1 & \mathbf{X}_2 & \dots & \mathbf{X}_n \end{bmatrix}$$

Points ($3 \times n$)

Cameras ($2m \times 3$)

- The measurement matrix $D = MS$ must have rank 3!

30

Computer Vision WS 14/15

C. Tomasi and T. Kanade, Shape and motion from image streams under orthography: A factorization method, IJCV, 9(2):137-154, November 1992. Slide credit: Svetlana Lazebnik B. Leibe

RWTH AACHEN UNIVERSITY

Factorizing the Measurement Matrix

Measurements = Motion \times Shape

$D = MS$

Computer Vision WS 14/15 31

Slide credit: Martial Hebert B. Leibe

RWTH AACHEN UNIVERSITY

Factorizing the Measurement Matrix

- Singular value decomposition of D:

Computer Vision WS 14/15 32

Slide credit: Martial Hebert

RWTH AACHEN UNIVERSITY

Factorizing the Measurement Matrix

- Singular value decomposition of D:

To reduce to rank 3, we just need to set all the singular values to 0 except for the first 3

Computer Vision WS 14/15 33

Slide credit: Martial Hebert

RWTH AACHEN UNIVERSITY

Factorizing the Measurement Matrix

- Obtaining a factorization from SVD:

Computer Vision WS 14/15 34

Slide credit: Martial Hebert

RWTH AACHEN UNIVERSITY

Factorizing the Measurement Matrix

- Obtaining a factorization from SVD:

Possible decomposition:
 $M = U_3 W_3^{1/2}$ $S = W_3^{1/2} V_3^T$

This decomposition minimizes $\|D - MS\|^2$

Computer Vision WS 14/15 35

Slide credit: Martial Hebert

RWTH AACHEN UNIVERSITY

Affine Ambiguity

- The decomposition is not unique. We get the same D by using any 3×3 matrix C and applying the transformations $M \rightarrow MC$, $S \rightarrow C^{-1}S$.
- That is because we have only an affine transformation and we have not enforced any Euclidean constraints (like forcing the image axes to be perpendicular, for example). We need a *Euclidean upgrade*.

Computer Vision WS 14/15 36

Slide credit: Martial Hebert B. Leibe

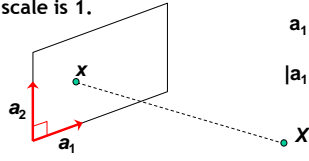
RWTH AACHEN UNIVERSITY

Estimating the Euclidean Upgrade

- Orthographic assumption: image axes are perpendicular and scale is 1.

$$\mathbf{a}_1 \cdot \mathbf{a}_2 = 0$$

$$|\mathbf{a}_1|^2 = |\mathbf{a}_2|^2 = 1$$



- This can be converted into a system of $3m$ equations:

$$\begin{cases} \hat{a}_{i1} \cdot \hat{a}_{i2} = 0 \\ |\hat{a}_{i1}| = 1 \\ |\hat{a}_{i2}| = 1 \end{cases} \Leftrightarrow \begin{cases} \mathbf{a}_{i1}^T \mathbf{C} \mathbf{C}^T \mathbf{a}_{i2} = 0 \\ \mathbf{a}_{i1}^T \mathbf{C} \mathbf{C}^T \mathbf{a}_{i1} = 1 \\ \mathbf{a}_{i2}^T \mathbf{C} \mathbf{C}^T \mathbf{a}_{i2} = 1 \end{cases}, \quad i = 1, \dots, m$$

for the transformation matrix $C \Rightarrow$ goal: estimate C

Computer Vision WS 14/15 37
Slide adapted from S. Lazebnik, M. Hebert, B. Leibe

RWTH AACHEN UNIVERSITY

Estimating the Euclidean Upgrade

- System of $3m$ equations:

$$\begin{cases} \hat{a}_{i1} \cdot \hat{a}_{i2} = 0 \\ |\hat{a}_{i1}| = 1 \\ |\hat{a}_{i2}| = 1 \end{cases} \Leftrightarrow \begin{cases} \mathbf{a}_{i1}^T \mathbf{C} \mathbf{C}^T \mathbf{a}_{i2} = 0 \\ \mathbf{a}_{i1}^T \mathbf{C} \mathbf{C}^T \mathbf{a}_{i1} = 1 \\ \mathbf{a}_{i2}^T \mathbf{C} \mathbf{C}^T \mathbf{a}_{i2} = 1 \end{cases}, \quad i = 1, \dots, m$$
- Let $L = \mathbf{C} \mathbf{C}^T$ $A_i = \begin{bmatrix} \mathbf{a}_{i1}^T \\ \mathbf{a}_{i2}^T \end{bmatrix}$, $i = 1, \dots, m$
- Then this translates to $3m$ equations in L

$$A_i L A_i^T = I, \quad i = 1, \dots, m$$
 - Solve for L
 - Recover C from L by Cholesky decomposition: $L = \mathbf{C} \mathbf{C}^T$
 - Update M and S : $M = \mathbf{C} \mathbf{C}$, $S = \mathbf{C}^{-1} S$

Computer Vision WS 14/15 38
Slide adapted from S. Lazebnik, M. Hebert, B. Leibe

RWTH AACHEN UNIVERSITY

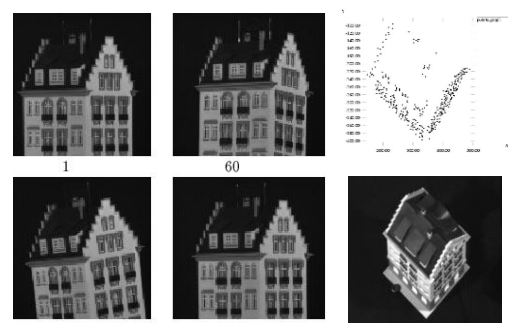
Algorithm Summary

- Given: m images and n features x_{ij}
- For each image i , center the feature coordinates.
- Construct a $2m \times n$ measurement matrix D :
 - Column j contains the projection of point j in all views
 - Row i contains one coordinate of the projections of all the n points in image i
- Factorize D :
 - Compute SVD: $D = U W V^T$
 - Create U_3 by taking the first 3 columns of U
 - Create V_3 by taking the first 3 columns of V
 - Create W_3 by taking the upper left 3×3 block of W
- Create the motion and shape matrices:
 - $M = U_3 W_3^{1/2}$ and $S = W_3^{1/2} V_3^T$ (or $M = U_3$ and $S = W_3 V_3^T$)
- Eliminate affine ambiguity

Computer Vision WS 14/15 39
Slide credit: Martial Hebert

RWTH AACHEN UNIVERSITY

Reconstruction Results

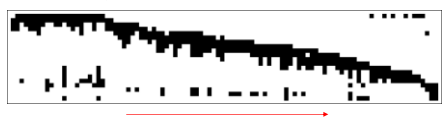


Computer Vision WS 14/15 40
C. Tomasi and T. Kanade, *Shape and motion from image streams under orthography: A factorization method*, *IJCV*, 9(2):137-154, November 1992.
Slide credit: Svetlana Lazebnik, B. Leibe, Image Source: Tomasi & Kanade

RWTH AACHEN UNIVERSITY

Dealing with Missing Data

- So far, we have assumed that all points are visible in all views
- In reality, the measurement matrix typically looks something like this:



Computer Vision WS 14/15 41
Slide credit: Svetlana Lazebnik, B. Leibe

RWTH AACHEN UNIVERSITY

Dealing with Missing Data

- Possible solution: decompose matrix into dense sub-blocks, factorize each sub-block, and fuse the results
 - Finding dense maximal sub-blocks of the matrix is NP-complete (equivalent to finding maximal cliques in a graph)
- Incremental bilinear refinement
 - Perform factorization on a dense sub-block

Computer Vision WS 14/15 42
F. Rothganger, S. Lazebnik, C. Schmid, and J. Ponce, *Segmenting, Modeling, and Matching Video Clips Containing Multiple Moving Objects*, *PAMI* 2007.
Slide credit: Svetlana Lazebnik

Computer Vision WS 14/15 RWTH AACHEN UNIVERSITY

Dealing with Missing Data

- Possible solution: decompose matrix into dense sub-blocks, factorize each sub-block, and fuse the results
 - Finding dense maximal sub-blocks of the matrix is NP-complete (equivalent to finding maximal cliques in a graph)
- Incremental bilinear refinement

- Perform factorization on a dense sub-block
- Solve for a new 3D point visible by at least two known cameras (linear least squares)

F. Rothganger, S. Lazebnik, C. Schmid, and J. Ponce. [Segmenting, Modeling, and Matching Video Clips Containing Multiple Moving Objects](#). PAMI 2007. 43

Slide credit: Svetlana Lazebnik

Computer Vision WS 14/15 RWTH AACHEN UNIVERSITY

Dealing with Missing Data

- Possible solution: decompose matrix into dense sub-blocks, factorize each sub-block, and fuse the results
 - Finding dense maximal sub-blocks of the matrix is NP-complete (equivalent to finding maximal cliques in a graph)
- Incremental bilinear refinement

- Perform factorization on a dense sub-block
- Solve for a new 3D point visible by at least two known cameras (linear least squares)
- Solve for a new camera that sees at least three known 3D points (linear least squares)

F. Rothganger, S. Lazebnik, C. Schmid, and J. Ponce. [Segmenting, Modeling, and Matching Video Clips Containing Multiple Moving Objects](#). PAMI 2007. 44

Slide credit: Svetlana Lazebnik

Computer Vision WS 14/15 RWTH AACHEN UNIVERSITY

Comments: Affine SfM

- Affine SfM was historically developed first.
- It is valid under the assumption of *affine cameras*.
 - Which does not hold for real physical cameras...
 - ...but which is still tolerable if the scene points are far away from the camera.
- For good results with real cameras, we typically need projective SfM.
 - Harder problem, more ambiguity
 - Math is a bit more involved... (Here, only basic ideas. If you want to implement it, please look at the H&Z book for details).

B. Leibe 45

Computer Vision WS 14/15 RWTH AACHEN UNIVERSITY

Topics of This Lecture

- Structure from Motion (SfM)
 - Motivation
 - Ambiguity
- Affine SfM
 - Affine cameras
 - Affine factorization
 - Euclidean upgrade
 - Dealing with missing data
- Projective SfM
 - Two-camera case
 - Projective factorization
 - Bundle adjustment
 - Practical considerations
- Applications

B. Leibe 46

Computer Vision WS 14/15 RWTH AACHEN UNIVERSITY

Projective Structure from Motion

- Given: m images of n fixed 3D points

$$x_{ij} = P_i X_j, \quad i = 1, \dots, m, \quad j = 1, \dots, n$$
- Problem: estimate m projection matrices P_i and n 3D points X_j from the mn correspondences x_{ij}

B. Leibe 47

Slide credit: Svetlana Lazebnik

Computer Vision WS 14/15 RWTH AACHEN UNIVERSITY

Projective Structure from Motion

- Given: m images of n fixed 3D points
 - $z_{ij} x_{ij} = P_i X_j, \quad i = 1, \dots, m, \quad j = 1, \dots, n$
- Problem: estimate m projection matrices P_i and n 3D points X_j from the mn correspondences x_{ij}
- With no calibration info, cameras and points can only be recovered up to a 4×4 projective transformation Q :

$$X \rightarrow QX, \quad P \rightarrow PQ^{-1}$$
- We can solve for structure and motion when

$$2mn \geq 11m + 3n - 15$$
- For two cameras, at least 7 points are needed.

B. Leibe 48

Slide credit: Svetlana Lazebnik

Computer Vision WS 14/15 RWTH AACHEN UNIVERSITY

Projective SfM: Two-Camera Case

- Assume fundamental matrix F between the two views
 - First camera matrix: $[I|0]Q^{-1}$
 - Second camera matrix: $[A|b]Q^{-1}$
- Let $\tilde{X} = QX$, then $z\tilde{x} = [I|0]\tilde{X}$, $z'\tilde{x}' = [A|b]\tilde{X}$
- And

$$z'\tilde{x}' = A[I|0]\tilde{X} + b = zAx + b$$

$$z'\tilde{x}' \times b = zAx \times b$$

$$(z'\tilde{x}' \times b) \cdot \tilde{x}' = (zAx \times b) \cdot \tilde{x}'$$

$$0 = (zAx \times b) \cdot \tilde{x}'$$
- So we have $\tilde{x}'^T [b_\times] Ax = 0$

$$F = [b_\times] A \quad b: \text{epipole } (F^T b = 0), \quad A = -[b_\times] F$$

Slide adapted from Svetlana Lazebnik B. Leibe FRP sec. 13.3.1 49

Computer Vision WS 14/15 RWTH AACHEN UNIVERSITY

Projective SfM: Two-Camera Case

- This means that if we can compute the fundamental matrix between two cameras, we can directly estimate the two projection matrices from F .
- Once we have the projection matrices, we can compute the 3D position of any point X by triangulation.
- How can we obtain both kinds of information at the same time?

B. Leibe 50

Computer Vision WS 14/15 RWTH AACHEN UNIVERSITY

Projective Factorization

$$D = \begin{bmatrix} z_{11}x_{11} & z_{12}x_{12} & \dots & z_{1n}x_{1n} \\ z_{21}x_{21} & z_{22}x_{22} & \dots & z_{2n}x_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ z_{m1}x_{m1} & z_{m2}x_{m2} & \dots & z_{mn}x_{mn} \end{bmatrix} = \begin{bmatrix} P_1 \\ P_2 \\ \vdots \\ P_m \end{bmatrix} \begin{bmatrix} X_1 & X_2 & \dots & X_n \end{bmatrix}$$

Points ($4 \times n$)
Cameras ($3m \times 4$)

$$D = MS \text{ has rank 4}$$

- If we knew the depths z , we could factorize D to estimate M and S .
- If we knew M and S , we could solve for z .
- Solution: iterative approach (alternate between above two steps).

Slide credit: Svetlana Lazebnik B. Leibe 51

Computer Vision WS 14/15 RWTH AACHEN UNIVERSITY

Sequential Structure from Motion

- Initialize motion from two images using fundamental matrix
- Initialize structure
- For each additional view:
 - Determine projection matrix of new camera using all the known 3D points that are visible in its image - *calibration*

Slide credit: Svetlana Lazebnik B. Leibe 52

Computer Vision WS 14/15 RWTH AACHEN UNIVERSITY

Sequential Structure from Motion

- Initialize motion from two images using fundamental matrix
- Initialize structure
- For each additional view:
 - Determine projection matrix of new camera using all the known 3D points that are visible in its image - *calibration*
 - Refine and extend structure: compute new 3D points, re-optimize existing points that are also seen by this camera - *triangulation*

Slide credit: Svetlana Lazebnik B. Leibe 53

Computer Vision WS 14/15 RWTH AACHEN UNIVERSITY

Sequential Structure from Motion

- Initialize motion from two images using fundamental matrix
- Initialize structure
- For each additional view:
 - Determine projection matrix of new camera using all the known 3D points that are visible in its image - *calibration*
 - Refine and extend structure: compute new 3D points, re-optimize existing points that are also seen by this camera - *triangulation*
- Refine structure and motion: *bundle adjustment*

Slide credit: Svetlana Lazebnik B. Leibe 54

RWTH AACHEN
UNIVERSITY

Bundle Adjustment

- Non-linear method for refining structure and motion
- Minimizing mean-square reprojection error

$$E(\mathbf{P}, \mathbf{X}) = \sum_{i=1}^m \sum_{j=1}^n D(\mathbf{x}_{ij}, \mathbf{P}_i \mathbf{X}_j)^2$$

Slide credit: Svetlana Lazebnik
B. Leibe

55

RWTH AACHEN
UNIVERSITY

Bundle Adjustment

- Seeks the Maximum Likelihood (ML) solution assuming the measurement noise is Gaussian.
- It involves adjusting the bundle of rays between each camera center and the set of 3D points.
- Bundle adjustment should generally be used as the final step of any multi-view reconstruction algorithm.
 - Considerably improves the results.
 - Allows assignment of individual covariances to each measurement.
- However...
 - It needs a good initialization.
 - It can become an extremely large minimization problem.
- Very efficient algorithms available.

Slide credit: Svetlana Lazebnik
B. Leibe

56

RWTH AACHEN
UNIVERSITY

Projective Ambiguity

- If we don't know anything about the camera or the scene, the best we can get with this is a reconstruction up to a projective ambiguity Q .
 - This can already be useful.
 - E.g. we can answer questions like "at what point does a line intersect a plane"?
- If we want to convert this to a "true" reconstruction, we need a **Euclidean upgrade**.
 - Need to put in additional knowledge about the camera (calibration) or about the scene (e.g. from markers).
 - Several methods available (see F&P Chapter 13.5 or H&Z Chapter 19)

Slide credit: Svetlana Lazebnik
B. Leibe

57

RWTH AACHEN
UNIVERSITY

Self-Calibration

- Self-calibration (auto-calibration) is the process of determining intrinsic camera parameters directly from uncalibrated images.
- For example, when the images are acquired by a single moving camera, we can use the constraint that the intrinsic parameter matrix remains fixed for all the images.
 - Compute initial projective reconstruction and find 3D projective transformation matrix Q such that all camera matrices are in the form $\mathbf{P}_i = \mathbf{K} [\mathbf{R}_i | \mathbf{t}_i]$.
- Can use constraints on the form of the calibration matrix: square pixels, zero skew, fixed focal length, etc.

Slide credit: Svetlana Lazebnik
B. Leibe

58

RWTH AACHEN
UNIVERSITY

Practical Considerations (1)

1. Role of the baseline
 - Small baseline: large depth error
 - Large baseline: difficult search problem
- Solution
 - Track features between frames until baseline is sufficient.

Slide adapted from Steve Seitz
B. Leibe

59

RWTH AACHEN
UNIVERSITY

Practical Considerations (2)

2. There will still be many outliers
 - Incorrect feature matches
 - Moving objects
 ⇒ Apply RANSAC to get robust estimates based on the inlier points.
3. Estimation quality depends on the point configuration
 - Points that are close together in the image produce less stable solutions.
 ⇒ Subdivide image into a grid and try to extract about the same number of features per grid cell.

Slide adapted from Steve Seitz
B. Leibe

60

RWTH AACHEN
UNIVERSITY

General Guidelines

- Use calibrated cameras wherever possible.
 - It makes life so much easier, especially for SfM.
- SfM with 2 cameras is *far* more robust than with a single camera.
 - Triangulate feature points in 3D using stereo.
 - Perform 2D-3D matching to recover the motion.
 - More robust to loss of scale (main problem of 1-camera SfM).
- Any constraint on the setup can be useful
 - E.g. square pixels, zero skew, fixed focal length in each camera
 - E.g. fixed baseline in stereo SfM setup
 - E.g. constrained camera motion on a ground plane
 - Making best use of those constraints may require adapting the algorithms (some known results are described in H&Z).

61

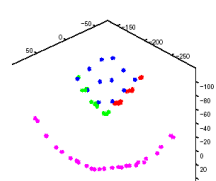
Computer Vision WS 14/15

B. Leibe

RWTH AACHEN
UNIVERSITY

Structure-from-Motion: Limitations

- Very difficult to reliably estimate metric SfM unless
 - Large (x or y) motion *or*
 - Large field-of-view and depth variation
- Camera calibration important for Euclidean reconstruction
- Need good feature tracker



62

Computer Vision WS 14/15

B. Leibe

Slide adapted from Steve Seitz

RWTH AACHEN
UNIVERSITY

Topics of This Lecture

- Structure from Motion (SfM)
 - Motivation
 - Ambiguity
- Affine SfM
 - Affine cameras
 - Affine factorization
 - Euclidean upgrade
 - Dealing with missing data
- Projective SfM
 - Two-camera case
 - Projective factorization
 - Bundle adjustment
 - Practical considerations
- Applications

63

Computer Vision WS 14/15

B. Leibe

RWTH AACHEN
UNIVERSITY

Commercial Software Packages

- boujou (<http://www.2d3.com/>)
- PFTrack (<http://www.thepixelfarm.co.uk/>)
- MatchMover (<http://www.realviz.com/>)
- SynthEyes (<http://www.ssontech.com/>)
- Icarus (<http://aig.cs.man.ac.uk/research/reveal/icarus/>)
- Voodoo Camera Tracker (<http://www.digilab.uni-hannover.de/>)


64

Computer Vision WS 14/15

B. Leibe

RWTH AACHEN
UNIVERSITY

Applications: Matchmoving



- Putting virtual objects into real-world videos
 - [Original sequence](#)
 - [Tracked features](#)
 - [SfM results](#)
 - [Final video](#)

66


Computer Vision WS 14/15

B. Leibe

Videos from Stefan Hafeneeger

RWTH AACHEN
UNIVERSITY

Applications: Large-Scale SfM from Flickr



67

Computer Vision WS 14/15

B. Leibe

S. Agarwal, N. Snavely, I. Simon, S.M. Seitz, R. Szeliski, [Building Rome in a Day](http://grail.cs.washington.edu/rome/), ICCV'09, 2009. (Video from <http://grail.cs.washington.edu/rome/>)

References and Further Reading

- A (relatively short) treatment of affine and projective SfM and the basic ideas and algorithms can be found in Chapters 12 and 13 of

D. Forsyth, J. Ponce,
Computer Vision - A Modern Approach,
Prentice Hall, 2003



- More detailed information (if you really want to implement this) and better explanations can be found in Chapters 10, 18 (factorization) and 19 (self-calibration) of

R. Hartley, A. Zisserman
Multiple View Geometry in Computer Vision
2nd Ed., Cambridge Univ. Press, 2004

