BMVC 2013
Incremental Line-based 3D Reconstruction using Geometric Constraints

Manuel Hofer, Andreas Wendel, Horst Bischof
Motivation

- Traditional Structure-from-Motion (SfM)
  - Using multiple images
  - Usually point based
  - Delivers accurate results for highly textured objects → many feature points
- Untextured scenes? (wiry objects, ...)

TU Graz | Aerial Vision Group
http://aerial.icg.tugraz.at
Motivation

• Alternative: Line-based 3D Reconstruction
  • Suitable for urban- and indoor scenes containing texture-less objects
  • Procedure similar to point-based methods:

<table>
<thead>
<tr>
<th></th>
<th>Points</th>
<th>Line-segments</th>
</tr>
</thead>
<tbody>
<tr>
<td>Feature detection</td>
<td>e.g. SIFT [Lowe, 2004]</td>
<td>e.g. LSD [Gioi et al., 2010]</td>
</tr>
<tr>
<td>Feature description + matching</td>
<td></td>
<td>e.g. MSLD [Zhiheng et al., 2009]</td>
</tr>
<tr>
<td>Pose estimation + reconstruction</td>
<td>e.g. [Irschara et al., 2010]</td>
<td>e.g. [Elqursh and Elgammal, 2011]</td>
</tr>
</tbody>
</table>
Motivation

- Alternative: Line-based 3D Reconstruction
  - Suitable for urban- and indoor scenes containing texture-less objects
  - Procedure similar to point-based methods:

<table>
<thead>
<tr>
<th></th>
<th>Points</th>
<th>Line-segments</th>
</tr>
</thead>
<tbody>
<tr>
<td>Feature detection</td>
<td>e.g. SIFT [Lowe, 2004]</td>
<td>e.g. LSD [Gioi et al., 2010]</td>
</tr>
<tr>
<td>Feature description +</td>
<td></td>
<td></td>
</tr>
<tr>
<td>matching</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pose estimation +</td>
<td>e.g. [Irschara et al., 2010]</td>
<td>e.g. [Elqursh and Elgammal, 2011]</td>
</tr>
<tr>
<td>reconstruction</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Line-segment Matching

- Usually appearance-based
  - Local descriptor based on gradient and color information from rectangular patch around the segment
    MSLD [Zhiheng et al., 2009], SILT [Khaleghi et al., 2009]
  - Color histograms along the line
    [Bay et al., 2005]
  - Does not work for wiry structures!
Line-segment Matching

- Usually appearance-based
  - Local descriptor based on gradient and color information from rectangular patch around the segment
    MSLD [Zhiheng et al., 2009], SILT [Khaleghi et al., 2009]
  - Color histograms along the line
    [Bay et al., 2005]
  - Does not work for wiry structures!
Appearance-less Approaches

- Jain et al., 2010
  - Assumes known cameras
  - Line-segments are not directly matched
  - Estimation of 3D line position:
    - Compute all possible locations in a certain sweeping range
    - Evaluate using multi-view backprojection and gradient scoring
    - Obtain final result and remove outliers by spatial clustering
  - Accurate results, but very time-consuming!
    (several hours for one image sequence, reported in the paper)
Appearance-less Approaches

- Hofer et al., 2013
  - Lines cannot be located at any 3D position
  - Use epipolar guided multi-view matching to compute discrete hypotheses set for each segment
  - Adapt gradient scoring and clustering from [Jain et al., 2010]
- Faster, but still slow...
Appearance-less Approaches

- Hofer et al., 2013
  - Lines cannot be located at any 3D position
  - Use epipolar guided multi-view matching to compute discrete hypotheses set for each segment
  - Adapt gradient scoring and clustering from [Jain et al., 2010]
  - Faster, but still slow…
Appearance-less Approaches

• Hofer et al., 2013
  • Lines cannot be located at any 3D position
  • Use epipolar guided multi-view matching to compute discrete hypotheses set for each segment
  • Adapt gradient scoring and clustering from [Jain et al., 2010]
  • Faster, but still slow…
Appearance-less Approaches

- Hofer et al., 2013
  - Lines cannot be located at any 3D position
  - Use epipolar guided multi-view matching to compute discrete hypotheses set for each segment
  - Adapt gradient scoring and clustering from [Jain et al., 2010]
  - Faster, but still slow...
Example Result

- Power Pylon (106 images)
  - Using [Hofer et al., 2013]
Example Result

- Power Pylon (106 images)
  - Using [Hofer et al., 2013]
  - Time: 67min (lines only...)
Remaining Challenges

- Mentioned algorithms deliver accurate results for the general case, but...
Remaining Challenges

- Mentioned algorithms deliver accurate results for the general case, but...

1. Cameras have to be known beforehand
   - Not useful for real-time applications (e.g. model-based tracking)
   - Is it possible to perform appearance-less 3D reconstruction online?
Remaining Challenges

• Mentioned algorithms deliver accurate results for the general case, but...

1. Cameras have to be known beforehand
   • Not useful for real-time applications (e.g. model-based tracking)
   • Is it possible to perform appearance-less 3D reconstruction online?

2. Very time-consuming
   • Bottlenecks are the gradient scoring and the clustering procedure at the end
   • Is it possible to avoid the scoring process at all and cluster corresponding hypotheses on-the-fly?
Remaining Challenges

• Mentioned algorithms deliver accurate results for the general case, but...

1. Cameras have to be known beforehand
   • Not useful for real-time applications (e.g. model-based tracking)
   • Is it possible to perform appearance-less 3D reconstruction online?

2. Very time-consuming
   • Bottlenecks are the gradient scoring and the clustering procedure at the end
   • Is it possible to avoid the scoring process at all and cluster corresponding hypotheses on-the-fly?

3. Reconstruction scale has to be known
   • Spatial clustering otherwise not possible
   • Is it possible to derive the clustering radius from the image space without knowing the exact reconstruction scale?
Incremental 3D Reconstruction

- We integrate the line reconstruction process into an online SfM system [Hoppe et al., 2012], to obtain live camera poses
- Uses SIFT feature matching [Lowe, 2004]
Incremental 3D Reconstruction

- We integrate the line reconstruction process into an online SfM system [Hoppe et al., 2012], to obtain live camera poses
- Uses SIFT feature matching [Lowe, 2004]
Incremental 3D Reconstruction

- We integrate the line reconstruction process into an online SfM system [Hoppe et al., 2012], to obtain live camera poses
- Uses SIFT feature matching [Lowe, 2004]
Incremental 3D Reconstruction

- We integrate the line reconstruction process into an online SfM system [Hoppe et al., 2012], to obtain live camera poses
- Uses SIFT feature matching [Lowe, 2004]
Incremental 3D Reconstruction

- Line matching procedure similar to [Hofer et al., 2013]
  - Epipolar guided matching
  - One 2D segment → several possible matches
- Instead of keeping one hypothesis per 2D segment, we keep all possible hypotheses until a decision can be made
  - Scene coverage may be still too small to decide which hypothesis is correct
- We perform on the fly grouping to cluster corresponding segments together
  - New line segments are added to existing hypotheses rather than creating new ones for each segment
  → new incremental result after each new image
Incremental Line-based 3D Reconstruction

Incremental 3D Reconstruction

- Reconstruction procedure:
  - Create an initial hypotheses set $H$
Incremental 3D Reconstruction

- Reconstruction procedure:
  - Create an initial hypotheses set $H$
Incremental 3D Reconstruction

- Reconstruction procedure:
  - Create an initial hypotheses set $H$

![Diagram showing line matching between two images $I_1$ and $I_2$.]
Incremental 3D Reconstruction

- Reconstruction procedure:
  - Create an initial hypotheses set $H$
Incremental 3D Reconstruction

- Reconstruction procedure:
  - Create an initial hypotheses set $H$
Incremental 3D Reconstruction

- Reconstruction procedure:
  - Create an initial hypotheses set $H$

find matching candidates
3D Line Segment Hypothesis

- Each hypothesis \( h \) consists of:
  - Triangulated line segment \( K_h \)
  - Set of corresponding 2D line segments \( L \), and cameras \( C \)
  - Score \( s(h) \) and corresponding camera \( C^*(h) \) defined as follows:

\[
s(h) = 1 - \min_{C_i \in C(h)} \left\{ \left| \frac{\vec{K}_h}{||\vec{K}_h||} \cdot \frac{\vec{C}_i}{||\vec{C}_i||} \right| \right\}, \quad C^*(h) = \arg\max_{C_i \in C(h)} s(h), \quad C_i \in C(h)
\]

→ score high for hypotheses with a large angle between the 3D line segment and one of the referenced cameras
Incremental 3D Reconstruction

- Reconstruction procedure:
  - Integrate new image
Incremental 3D Reconstruction

- Reconstruction procedure:
  - Integrate new image
Incremental 3D Reconstruction

- Reconstruction procedure:
  - Integrate new image
Incremental 3D Reconstruction

- Reconstruction procedure:
  - Integrate new image
Incremental 3D Reconstruction

- Reconstruction procedure:
  - Integrate new image
Incremental 3D Reconstruction

- Reconstruction procedure:
  - Integrate new image

check existing hypotheses!

find matching candidates

matching

triangulate

matching
Incremental 3D Reconstruction

- Reconstruction procedure:
  - Integrate new image

\[ \text{matching} \]

\[ \text{match!} \]

\[ \text{triangulate} \]

\[ \text{no match!} \]

\[ \text{check existing hypotheses!} \]
Matching Constraints

• When do we add a new line segment to an existing hypothesis?
  • If distance in 3D is lower than $r$ and distance in image space is lower than $\sigma$ (→ backprojection)

• How to define these thresholds?
  • $r$ requires scale information...
  • $\sigma$ can be chosen more easily (e.g. 1px)
Matching Constraints

• When do we add a new line segment to an existing hypothesis?
  • If distance in 3D is lower than $r$ and distance in image space is lower than $\sigma$ (→ backprojection)

• How to define these thresholds?
  • $r$ requires scale information...
  • $\sigma$ can be chosen more easily (e.g. 1px)

• Deriving $r$ from $\sigma$:
  • Project hypotheses $h \in H$ back into corresponding images
Matching Constraints

- When do we add a new line segment to an existing hypothesis?
  - If distance in 3D is lower than $r$ and distance in image space is lower than $\sigma$ (→ backprojection)

- How to define these thresholds?
  - $r$ requires scale information...
  - $\sigma$ can be chosen more easily (e.g. 1px)

• Deriving $r$ from $\sigma$:
  - Project hypotheses $h \in H$ back into corresponding images
  - Shift by $\sigma$ in same orthogonal direction and triangulate as $h'$
Matching Constraints

• When do we add a new line segment to an existing hypothesis?
  • If distance in 3D is lower than $r$ and distance in image space is lower than $\sigma$ (→ backprojection)

• How to define these thresholds?
  • $r$ requires scale information...
  • $\sigma$ can be chosen more easily (e.g. 1px)

• Deriving $r$ from $\sigma$:
  • Project hypotheses $h \in H$ back into corresponding images
  • Shift by $\sigma$ in same orthogonal direction and triangulate as $h'$
Matching Constraints

• When do we add a new line segment to an existing hypothesis?
  • If distance in 3D is lower than $r$ and distance in image space is lower than $\sigma$ (→ backprojection)

• How to define these thresholds?
  • $r$ requires scale information...
  • $\sigma$ can be chosen more easily (e.g. 1px)

• Deriving $r$ from $\sigma$:
  • Project hypotheses $h \in H$ back into corresponding images
  • Shift by $\sigma$ in same orthogonal direction and triangulate as $h'$
  • $r$ is the distance between the original and the shifted 3D segment
Matching Constraints

• When do we add a new line segment to an existing hypothesis?
  • If distance in 3D is lower than \( r \) and distance in image space is lower than \( \sigma \) (\( \rightarrow \) backprojection)

• How to define these thresholds?
  • \( r \) requires scale information...
  • \( \sigma \) can be chosen more easily (e.g. 1px)

• Deriving \( r \) from \( \sigma \):
  • Project hypotheses \( h \in H \) back into corresponding images
  • Shift by \( \sigma \) in same orthogonal direction and triangulate as \( h' \)
  • \( r \) is the distance between the original and the shifted 3D segment
  • For robustness, compute characteristic \( r(C) \) for each camera (median of referenced hypotheses)
  • Use \( r(C^*(h)) \) for further matching procedures involving \( h \)
Matching Constraints

• When do we add a new line segment to an existing hypothesis?
  • If distance in 3D is lower than $r$ and distance in image space is lower than $\sigma$ (→ backprojection)

• How to define these thresholds?
  • $r$ requires scale information...
  • $\sigma$ can be chosen more easily (e.g. 1px)

• Deriving $r$ from $\sigma$:
  • Project hypotheses $h \in H$ back into corresponding images
  • Shift by $\sigma$ in same orthogonal direction and triangulate as $h'$
  • $r$ is the distance between the original and the shifted 3D segment
  • For robustness, compute characteristic $r(C)$ for each camera (median of referenced hypotheses)
  • Use $r(C^*(h))$ for further matching procedures involving $h$

→ No dependence on reconstruction scale!
Incremental Results

- Simple greedy algorithm:
  - Sort current hypotheses set $H$ by number of participating line segments (hypothesis size)
  - If equal, sort by reprojection error
  - Iterate over sorted set:
    - If hypothesis size $\geq \lambda$ and $s(h) > 0.5$ $\rightarrow$ inlier [all other hypotheses referenced by any segment in $h$ are considered to be outliers and skipped (not erased!)]
    - Else $\rightarrow$ outlier
  - Remove unpromising hypotheses to prevent performance break-down
Incremental Results

- Simple greedy algorithm:
  - Sort current hypotheses set $H$ by number of participating line segments (hypothesis size)
  - If equal, sort by reprojection error
  - Iterate over sorted set:
    - If hypothesis size $\geq \lambda$ and $s(h) > 0.5 \rightarrow$ inlier
      [all other hypotheses referenced by any segment in $h$ are considered to be outliers and skipped (not erased!)]
    - Else $\rightarrow$ outlier
  - Remove unpromising hypotheses to prevent performance break-down

$\rightarrow$ Purely geometric hypothesis verification! No gradient scoring necessary!
Comparison: Offline vs. Online

- Pylon Sequence:
  - 106 ground-level images
Comparison: Offline vs. Online

- **Pylon Sequence:**
  - 106 ground-level images

**Offline**
Runtime: 67 minutes (lines only)

**Online**
Comparison: Offline vs. Online

- Pylon Sequence:
  - 106 ground-level images

Offline
Runtime: 67 minutes (lines only)

Online
Runtime: 9 minutes (incl. SfM)
Evaluation

- Timber-frame Sequence
  - Synthetic sequence (240 images)
  - Evaluation in terms of root mean square (RMS) error compared to ground truth CAD model

Jain et al., 2010
RMSE: 0.291

Hofer et al., 2013
RMSE: 0.094

Hofer et al., 2013a
RMSE: 0.196
Evaluation

- Timber-frame Sequence
  - Synthetic sequence (240 images)
  - Evaluation in terms of root mean square (RMS) error compared to ground truth CAD model

- Jain et al., 2010
  - RMSE: 0.291
  - Runtime: several hours...

- Hofer et al., 2013
  - RMSE: 0.094

- Hofer et al., 2013a
  - RMSE: 0.196
Evaluation

- Timber-frame Sequence
  - Synthetic sequence (240 images)
  - Evaluation in terms of root mean square (RMS) error compared to ground truth CAD model

- Jain et al., 2010
  - RMSE: 0.291
  - Runtime: several hours...

- Hofer et al., 2013
  - RMSE: 0.094
  - Runtime: 45 minutes

- Hofer et al., 2013a
  - RMSE: 0.196
Evaluation

- Timber-frame Sequence
  - Synthetic sequence (240 images)
  - Evaluation in terms of root mean square (RMS) error compared to ground truth CAD model

Jain et al., 2010
RMSE: 0.291
Runtime: several hours...

Hofer et al., 2013
RMSE: 0.094
Runtime: 45 minutes

Hofer et al., 2013a
RMSE: 0.196
Runtime: 12 minutes
Conclusion
Conclusion

1. Is it possible to perform appearance-less 3D reconstruction online?
Conclusion

1. **Is it possible to perform appearance-less 3D reconstruction online?**
   - Yes, it is possible to extend the principles introduced by [Jain et al., 2010] and [Hofer et al., 2013] for online processing!
Conclusion

1. Is it possible to perform appearance-less 3D reconstruction online?
   - Yes, it is possible to extend the principles introduced by [Jain et al., 2010] and [Hofer et al., 2013] for online processing!

2. Is it possible to avoid the scoring process at all and cluster corresponding hypotheses on-the-fly?
1. Is it possible to perform appearance-less 3D reconstruction online?
   • Yes, it is possible to extend the principles introduced by [Jain et al., 2010] and [Hofer et al., 2013] for online processing!

2. Is it possible to avoid the scoring process at all and cluster corresponding hypotheses on-the-fly?
   • Yes, we can group corresponding line segments together on-the-fly and verify them through their cluster size, without the need for backprojection and gradient scoring.

Conclusion
Conclusion

1. Is it possible to perform appearance-less 3D reconstruction online?
   - Yes, it is possible to extend the principles introduced by [Jain et al., 2010] and [Hofer et al., 2013] for online processing!

2. Is it possible to avoid the scoring process at all and cluster corresponding hypotheses on-the-fly?
   - Yes, we can group corresponding line segments together on-the-fly and verify them through their cluster size, without the need for backprojection and gradient scoring.

3. Is it possible to derive the clustering radius from the image space without knowing the exact reconstruction scale?
Conclusion

1. **Is it possible to perform appearance-less 3D reconstruction online?**
   - Yes, it is possible to extend the principles introduced by [Jain et al., 2010] and [Hofer et al., 2013] for online processing!

2. **Is it possible to avoid the scoring process at all and cluster corresponding hypotheses on-the-fly?**
   - Yes, we can group corresponding line segments together on-the-fly and verify them through their cluster size, without the need for backprojection and gradient scoring.

3. **Is it possible to derive the clustering radius from the image space without knowing the exact reconstruction scale?**
   - Yes, it is possible to derive the clustering radius directly from the image space using a pre-defined maximum uncertainty $\sigma$. 
Thank you for your attention!

More information available at http://aerial.icg.tugraz.at

This work has been supported by the Austrian Research Promotion Agency (FFG) project FIT-IT Pegasus (825841/10397) and OMICRON electronics GmbH.
References

- [Jain et al., 2010] A. Jain, C. Kurz, T. Thormaehlen, and H. Seidel
  *Exploiting global connectivity constraints for reconstruction of 3D line segments from images*, CVPR, 2010

- [Zhiheng et al., 2009] W. Zhiheng, W. Fuchao, and H. Zhanyi
  *MSLD: A robust descriptor for line matching*, Pattern Recognition, 2009

- [Khaleghi et al., 2009] B. Khaleghi, M. Baklouti, and F. Karray
  *SILT: Scale-invariant line transform*, CIRA, 2009

- [Hofer et al., 2013] M. Hofer, A. Wendel, and H. Bischof
  *Line-based 3D Reconstruction of Wiry Objects*, CVWW, 2013

- [Hofer et al., 2013a] M. Hofer, A. Wendel, and H. Bischof

- [Hoppe et al., 2012] C. Hoppe, M. Klopschitz, M. Rumpler, A. Wendel, S. Kluckner, H. Bischof, and G. Reitmayr
  *Online Feedback for Structure-from-Motion Image Acquisition*, BMVC, 2012

  *LSD: A fast line segment detector with a false detection control*, PAMI, 2010

  *Distinctive Image Features from Scale-Invariant Keypoints*, IJCV, 2004

- [Irschara et al., 2009] A. Irschara, V. Kaufmann, M. Klopschitz, H. Bischof, and F. Leberl
  *Towards fully automatic photogrammetric reconstruction using digital images taken from UAVs*, ISPRS, 2010


- [Bay et al., 2005] H. Bay, V. Ferrari, and L. Van Gool
  *Wide-baseline stereo matching with line segments*, CVPR, 2005