Camera Self-Calibration from Tracking of Moving Persons

ICPR 2016 Oral Session MoAT3

Zheng Tang\textsuperscript{1}, Yen-Shuo Lin\textsuperscript{2}, Kuan-Hui Lee\textsuperscript{1}, Jenq-Neng Hwang\textsuperscript{1}, Jen-Hui Chuang\textsuperscript{2}, Zhijun Fang\textsuperscript{3}

\textsuperscript{1} University of Washington, United States;
\textsuperscript{2} National Chiao Tung University, Taiwan;
\textsuperscript{3} Shanghai University of Engineering Science, China
Agenda

• Introduction
• System Overview
• Computation of Vanishing Points
  – Object Tracking and Head/Foot Localization
  – Vanishing Points Estimation
• Self-Calibration by Optimization
• Experimental Results
• Conclusion
Agenda

• Introduction

• System Overview

• Computation of Vanishing Points
  – Object Tracking and Head/Foot Localization
  – Vanishing Points Estimation

• Self-Calibration by Optimization

• Experimental Results

• Conclusion
Introduction

- Geometry of camera model

\[ [u, v, 1]^T \sim P \cdot [X, Y, Z, 1]^T \]

\[ P = K \cdot [R|t] \]

\[ K = \begin{bmatrix} f_x & s & c_x \\ 0 & f_y & c_y \\ 0 & 0 & 1 \end{bmatrix} \quad t = \begin{bmatrix} t_X \\ t_Y \\ t_Z \end{bmatrix} \]

\[ R = R_Z R_X R_Y \]

\[ R_Z = \begin{bmatrix} \cos(\text{roll}) & -\sin(\text{roll}) & 0 \\ \sin(\text{roll}) & \cos(\text{roll}) & 0 \\ 0 & 0 & 1 \end{bmatrix} \]

\[ R_X = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos(\text{pitch}) & -\sin(\text{pitch}) \\ 0 & \sin(\text{pitch}) & \cos(\text{pitch}) \end{bmatrix} \]

\[ R_Y = \begin{bmatrix} \cos(\text{yaw}) & 0 & -\sin(\text{yaw}) \\ 0 & 1 & 0 \\ \sin(\text{yaw}) & 0 & \cos(\text{yaw}) \end{bmatrix} \]
Introduction

- Categories of camera calibration
  - Calibration using calibrated templates
  - Self-calibration/auto-calibration
    - Static scene structures
    - Object motion -> Moving-person tracking
Introduction

• Self-calibration from tracking of moving persons [Lv et al., 2002]

\[
\text{roll} = \tan^{-1}\left(\frac{v_{VZ} - v_{VX}}{u_{VX} - u_{VZ}}\right)
\]

\[
f_X = f_Y = \sqrt{-\left(v_{VX}^{\text{rot}} \cdot v_{VZ}^{\text{rot}} + u_{VX}^{\text{rot}} \cdot u_{VZ}^{\text{rot}}\right)}
\]

where \(v_{VX}^{\text{rot}} = \cos(\text{roll}) (v_p - v_{VX}) - \sin(\text{roll}) (u_{VX} - u_p),\)

\(v_{VZ}^{\text{rot}} = \cos(\text{roll}) (v_p - v_{VZ}) - \sin(\text{roll}) (u_{VZ} - u_p),\)

\(u_{VX}^{\text{rot}} = \cos(\text{roll}) (u_{VX} - u_p) + \sin(\text{roll})(v_p - v_{VX}),\)

and \(u_{VZ}^{\text{rot}} = \cos(\text{roll}) (u_{VZ} - u_p) + \sin(\text{roll})(v_p - v_{VZ}).\)

\[
pitch = \tan^{-1}\left(\frac{v_{VX}^{\text{rot}}}{f_X}\right)
\]

\[
yaw = -\tan^{-1}\left(\frac{f_X}{\cos(pitch) \cdot u_{VX}^{\text{rot}}}\right)
\]

Original assumptions:
(1) Central principal point
(2) Unit aspect ratio
(3) Zero skew

Challenges:
(1) How to find the accurate \(V_Y\) and \(L_H\)?
(2) How can we optimize all camera parameters (relax original assumptions)?
Agenda

• Introduction
• **System Overview**
• Computation of Vanishing Points
  – Object Tracking and Head/Foot Localization
  – Vanishing Points Estimation
• Self-Calibration by Optimization
• Experimental Results
• Conclusion
System Overview

- Input Video Sequence
  - Tracking and Segmentation Based on MAST
    - Head/Foot Localization
      - $V_Y$ Estimation Based on Mean Shift Clustering
      - $L_H$ Estimation Based on Laplace Linear Regression
  - $V_X$ and $V_Z$ Estimation
    - Initialization of Camera Parameters
  - Optimization of Camera Parameters by EDA
    - Output Projection Matrix from 3-D to 2-D
Agenda

• Introduction
• System Overview
• **Computation of Vanishing Points**
  – Object Tracking and Head/Foot Localization
  – Vanishing Points Estimation
• Self-Calibration by Optimization
• Experimental Results
• Conclusion
System Overview

Input Video Sequence

- Tracking and Segmentation Based on MAST
- Head/Foot Localization
  - $V_Y$ Estimation Based on Mean Shift Clustering
  - $L_H$ Estimation Based on Laplace Linear Regression
  - $V_X$ and $V_Z$ Estimation
  - Initialization of Camera Parameters
  - Optimization of Camera Parameters by EDA

Output Projection Matrix from 3-D to 2-D
Object Tracking and Segmentation Based on MAST

Multiple-kernel Adaptive Segmentation and Tracking

- Object merging

![Object Tracking and Segmentation Diagram]
Object Tracking and Segmentation Based on MAST

**Multiple-kernel Adaptive Segmentation and Tracking**

- Motivation

[Image: Diagram showing segmentation and tracking results with red and blue areas, and a black background.]
Object Tracking and Segmentation Based on MAST

Multiple-kernel Adaptive Segmentation and Tracking \cite{Tang et al., 2016}

- Change Detection / Segmentation
  - Change Detection Based on \textsc{SuBSENSE}
    - Including Shadow Detection Based on YCbCr Color Space
  - Regularization / Post-processing
  - Tracking by Kalman Filtering and Multiple Kernels Tracking
  - Bounding Box Change Restriction
    - Constructing YCbCr Kernel Histogram
    - Computing Color Similarity
    - Penalizing Thresholds and Expanding Kernel Region
    - Constructing CbCr Kernel Histogram
    - Computing Chromaticity Similarity
    - Penalizing Thresholds and Expanding Kernel Region

- Tracking

- Penalty Computation for Feedback

Input Frame

Output Tracking Result after Feedback Loop
Head/Foot Localization
Agenda

• Introduction
• System Overview
• Computation of Vanishing Points
  – Object Tracking and Head/Foot Localization
  – Vanishing Points Estimation
• Self-Calibration by Optimization
• Experimental Results
• Conclusion
System Overview

Input Video Sequence

Tracking and Segmentation Based on MAST

Head/Foot Localization

$V_Y$ Estimation Based on Mean Shift Clustering

$L_H$ Estimation Based on Laplace Linear Regression

$V_X$ and $V_Z$ Estimation

Initialization of Camera Parameters

Optimization of Camera Parameters by EDA

Output Projection Matrix from 3-D to 2-D
Vertical Vanishing Point ($V_Y$) Estimation
Based on Mean Shift Clustering

• Disadvantage of RANSAC
  – Failure when the number of outliers is significantly large

• Proposed method
  – Mean shift clustering among all the candidate points of $V_Y$
  – Choosing the mean point of the largest cluster as the estimated $V_Y$
Horizon Line ($L_H$) Estimation Based on Laplace Linear Regression

- **Disadvantage of RANSAC**
  - Setting of threshold parameter for inliers

- **Proposed method**
  - Formulating as convex optimization by Laplace linear regression

\[
p(y|x, w) = \text{Laplace}(y|w^T x) \propto \exp(-|y - w^T x|)
\]

\[
\min_{w, r} \sum_i r_i = \min_{w, r^+, r^-} \sum_i (r_i^+ + r_i^-)
\]

s. t. \( r_i^+ \geq 0, r_i^- \geq 0, w^T x_i + r_i^+ - r_i^- = y_i \)

\[
\min_{\theta} f^T \theta \text{ s. t. } A\theta \leq b, A_{eq} \theta = b_{eq}, l \leq \theta \leq u
\]

in which \( \theta = (w, r^+, r^-), f = [0, 1, 1], A = [], b = [], A_{eq} = [x, l, -l], b_{eq} = y, l = [-\infty, 0, 0] \) and \( u = [] \).
Agenda

• Introduction
• System Overview
• Computation of Vanishing Points
  – Object Tracking and Head/Foot Localization
  – Vanishing Points Estimation
• **Self-Calibration by Optimization**
• Experimental Results
• Conclusion
System Overview

Input Video Sequence

Tracking and Segmentation Based on MAST

Head/Foot Localization

$V_Y$ Estimation Based on Mean Shift Clustering

$L_H$ Estimation Based on Laplace Linear Regression

$V_X$ and $V_Z$ Estimation

Initialization of Camera Parameters

Optimization of Camera Parameters by EDA

Output Projection Matrix from 3-D to 2-D
Consider $\arg \min_x f(x)$

1. Randomly generate $R$ samples.
2. Calculate $f(x_i)$ of each sample, and sort the results.
3. Use the best $N$ results to generate a pdf with normal distribution.
4. If stopping criterion is not met, use the pdf to generate new $R$ samples, jump to 2.

In this example, $R = 12$, $N = 6$

until stopping criterion is met
Optimization of Camera Parameters by EDA

Estimation of Multivariate Normal Algorithm–global [Larrauaga et al., 2002]

- Extending from univariate EDA, for multivariate scenario (8 parameters)
  - Using multivariate normal density function as pdf
- Each projection matrix formed by a set of camera parameters is regarded as a sample.
- Using reprojection error on the ground plane as the evaluation
- Stopping criterion
  - Change of reprojection error between generations is small enough
  - Number of generations is too large
Advantages of the Proposed Formulation

• Optimizing all camera parameters simultaneously by EMNA_global

• Relaxing original assumptions on intrinsic camera parameters by allowing them to be optimized within given ranges

• Advantages of EDA [Hauschild et al., 2011]
  – Ability to adapt their operators to the structure of the problem
  – Prior knowledge exploitation
  – Reduced memory requirements
  – Implementation of parallel computation
Agenda

• Introduction
• System Overview
• Computation of Vanishing Points
  – Object Tracking and Head/Foot Localization
  – Vanishing Points Estimation
• Self-Calibration by Optimization
• **Experimental Results**
• Conclusion
Experimental Results

• 3 captured video sequences
  – Length: ~ 1 min 30 sec
  – Resolution: 640 * 480
  – Frame rate: 10 fps
  – Ground truth: Extracted using linear method based on 52, 52, and 38 measured 3-D points

• 1 video sequence from EPFL dataset
  – Length: 3 min
  – Resolution: 360 * 288
  – Frame rate: 25 fps
  – Ground truth: Extracted using Tsai’s method
## Experimental Results

<table>
<thead>
<tr>
<th>Seq. #</th>
<th>$f_x$ (pix.)</th>
<th>$f_y$ (pix.)</th>
<th>$c_x$ (pix.)</th>
<th>$c_y$ (pix.)</th>
<th>roll (deg.)</th>
<th>pitch (deg.)</th>
<th>yaw (deg.)</th>
<th>$\mu_{err}$ (pix.)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Ground Truth</td>
<td>731.3880</td>
<td>728.2518</td>
<td>322.1298</td>
<td>237.2676</td>
<td>-3.1371</td>
<td>16.2676</td>
<td>-78.3065</td>
<td>N/A</td>
</tr>
<tr>
<td>1. Proposed w/o EDA</td>
<td><strong>738.7650</strong></td>
<td><strong>738.7650</strong></td>
<td><strong>320.0000</strong></td>
<td><strong>240.0000</strong></td>
<td><strong>5.0689</strong></td>
<td><strong>17.6076</strong></td>
<td><strong>79.0154</strong></td>
<td><strong>2.50E-5</strong></td>
</tr>
<tr>
<td>1. Proposed</td>
<td>730.9167</td>
<td>735.9371</td>
<td>322.9955</td>
<td>236.1948</td>
<td>-5.0345</td>
<td>17.4224</td>
<td>-79.1491</td>
<td><strong>3.12E-5</strong></td>
</tr>
<tr>
<td>2. Ground Truth</td>
<td>731.3880</td>
<td>728.2518</td>
<td>322.1298</td>
<td>237.2676</td>
<td>-1.8887</td>
<td>11.0081</td>
<td>-68.7126</td>
<td>N/A</td>
</tr>
<tr>
<td>2. Proposed w/o EDA</td>
<td><strong>679.6617</strong></td>
<td><strong>679.6617</strong></td>
<td><strong>320.0000</strong></td>
<td><strong>240.0000</strong></td>
<td><strong>1.7928</strong></td>
<td><strong>10.7818</strong></td>
<td><strong>70.3027</strong></td>
<td><strong>4.6445</strong></td>
</tr>
<tr>
<td>3. Ground Truth</td>
<td>731.3880</td>
<td>728.2518</td>
<td>322.1298</td>
<td>237.2676</td>
<td>-0.3459</td>
<td>18.3846</td>
<td>-63.8778</td>
<td>N/A</td>
</tr>
<tr>
<td>3. Method in [10]</td>
<td>662.9474</td>
<td>662.9474</td>
<td>320.0000</td>
<td>240.0000</td>
<td>-0.2164</td>
<td>22.4663</td>
<td>-57.6830</td>
<td>0.5403</td>
</tr>
<tr>
<td>3. Proposed w/o EDA</td>
<td><strong>719.8882</strong></td>
<td><strong>719.8882</strong></td>
<td><strong>320.0000</strong></td>
<td><strong>240.0000</strong></td>
<td><strong>0.2693</strong></td>
<td><strong>17.4219</strong></td>
<td><strong>64.7125</strong></td>
<td><strong>0.3398</strong></td>
</tr>
<tr>
<td>3. Proposed</td>
<td>720.6649</td>
<td>729.5090</td>
<td>319.8556</td>
<td>240.6065</td>
<td>-0.2658</td>
<td>17.2493</td>
<td>-64.7081</td>
<td><strong>1.17E-4</strong></td>
</tr>
<tr>
<td>4. Ground Truth</td>
<td>437.2689</td>
<td>437.8792</td>
<td>173.7693</td>
<td>142.7878</td>
<td>1.5466</td>
<td>14.1153</td>
<td>-54.5257</td>
<td>N/A</td>
</tr>
<tr>
<td>4. Method in [6]</td>
<td>406.8041</td>
<td>406.8041</td>
<td>180.0000</td>
<td>144.0000</td>
<td>-0.2633</td>
<td>22.4482</td>
<td>-63.5813</td>
<td>0.5051</td>
</tr>
<tr>
<td>4. Method in [10]</td>
<td><strong>432.0973</strong></td>
<td><strong>432.0973</strong></td>
<td><strong>180.0000</strong></td>
<td><strong>144.0000</strong></td>
<td><strong>-0.2062</strong></td>
<td><strong>20.8494</strong></td>
<td><strong>45.6322</strong></td>
<td><strong>0.4321</strong></td>
</tr>
<tr>
<td>4. Proposed w/o EDA</td>
<td><strong>440.5366</strong></td>
<td><strong>440.5366</strong></td>
<td><strong>180.0000</strong></td>
<td><strong>144.0000</strong></td>
<td><strong>-0.4297</strong></td>
<td><strong>16.2182</strong></td>
<td><strong>55.8775</strong></td>
<td><strong>0.1858</strong></td>
</tr>
<tr>
<td>4. Proposed</td>
<td>442.4795</td>
<td>440.9664</td>
<td>176.2516</td>
<td>142.1498</td>
<td>0.4313</td>
<td>15.9846</td>
<td>-55.6434</td>
<td><strong>2.74E-5</strong></td>
</tr>
</tbody>
</table>
Experimental Results
Agenda

• Introduction
• System Overview
• Computation of Vanishing Points
  – Object Tracking and Head/Foot Localization
  – Vanishing Points Estimation
• Self-Calibration by Optimization
• Experimental Results
• Conclusion
Conclusion

• We proposed a robust single camera self-calibration method based on moving persons tracking.

• **Contribution (1):** Combining the state-of-the-art change detection (**SuBSENSE**) and tracking (**MAST**) to generate accurate head/foot localization

• **Contribution (2):** Introducing mean shift clustering and Laplace linear regression to the estimation of vanishing points

• **Contribution (3):** formulating the problem of camera parameters optimization by **EDA** that can relax the assumptions on unknown intrinsic parameters.
Thank you!

Q&A