Multi-Camera Rectification using Linearized Trifocal Tensor

F. Zilly\textsuperscript{1,2}, C. Riechert\textsuperscript{1}, M. Müller\textsuperscript{1}, W. Waizenegger\textsuperscript{1}, T. Sikora\textsuperscript{2}, P. Kaufl\textsuperscript{1}

\textsuperscript{1}Fraunhofer Institute for Telecommunications, Berlin, Germany
\textsuperscript{2}Technische Universität Berlin, Germany
frederik.zilly@hhi.fraunhofer.de

Abstract

Multi-camera systems such as linear camera arrays are commonly used to capture content for multi-baseline stereo estimation, view generation for autostereoscopic displays, or similar tasks. However, even after a careful mechanical alignment, residual vertical disparities and horizontal disparity offsets impair further processing steps. In consequence, the multi-camera content needs to be rectified on a common baseline. The trifocal tensor represents the geometry between three cameras and hence is a helpful tool to calibrate a multi-camera system, and to derive rectifying homographies. Against this background we propose a new method for a robust estimation of the trifocal tensor specialized for linear camera arrays and subsequent rectifying homography computation based on feature point triplets.

1. Introduction

Linear camera arrays have a long tradition in computer vision. A standard application is the generation of depth maps which can be estimated in a more precise and robust way than by using a stereo vision system only. A prominent example is the Multi-baseline stereo approach proposed in [13] where disparities measured between different camera pairs on the same baseline are mapped into an inverse-of-a-distance measure, to allow comparing and merging them. One crucial prerequisite of this approach is the absence of all vertical disparities among all camera images, and, that all horizontal disparities are proportional to each other, the proportionality factor being the ratio of the camera baselines. In the following, we will refer to the first requirement as vertical alignment, while calling the latter the horizontal alignment. Consequently, the aim of the multi-camera rectification algorithm is to ensure that these requirements are met.

Several trifocal or trinocular rectification algorithms are known in the literature. Trinocular rectification techniques which require calibrated cameras or dedicated calibration pattern have been proposed in [1,2,4,5,9]. In contrast, the method proposed by [15] works with uncalibrated cameras by using the trilinear tensor in its representation used in [14]. Other calibration-free approaches for trifocal rectification were proposed in [8,3,17]. However, these techniques target L-shaped trifocal setups while our focus lies on linear-camera arrays. A rectification method for three cameras in a horizontal setup using uncalibrated cameras has been proposed by [10] and extended towards four or more cameras by [16]. Both methods are based on [11]. Recently, Nozick extended [4] towards uncalibrated cameras [12]. A simple, though real-time capable multi-camera rectification algorithm which considers only vertical pixel shifting has been proposed in [7]. Apart from [16] which applies also a simple horizontal pixel shifting to achieve equidistant disparities, we do not know of other algorithms which target horizontal alignment and use uncalibrated cameras in a linear array. However, parallel epipolar lines are not a sufficient constraint for a proper horizontal alignment, as a stretching or offset of the horizontal disparities can still occur due to a horizontal shift of the principal point or a deviation of the pixel aspect ratio from 1.

Against this background we propose an algorithm which achieves vertical alignment, horizontal alignment, i.e. proportional horizontal disparities after rectification, and extracts this proportionality constant, i.e. the ratio of the camera baselines which do not need to be equidistant. We summarize the main properties of the proposed rectification method:

- Based on feature point triplets,
- Suitable for uncalibrated cameras,
- Elimination of vertical disparities,
Horizontal disparities become proportional to the ratio of the two camera baselines,
Provision of the proportionality constant $\beta$,
Linear estimation of the trifocal tensor,
Robust against noise and outliers.

We propose a rectification technique which is based on estimating the trifocal tensor which has been developed in a Taylor serie around the ideal state of perfectly aligned cameras.

2. Linearization of the Trifocal Tensor

The basic concept of the linearization of the trifocal tensor is a Taylor expansion of the tensor around a known state, which is in our example a camera configuration of a perfectly aligned linear camera array. The linearization approach similar to [18] allows us to compute the trifocal tensor by solving a linear set of equations. More importantly, once the result vector, i.e. the set of geometric parameters is known, it gives us direct access not only to the components of the trifocal tensor but also to the underlying geometry. These parameters can directly be used for the composition of rectifying homographies, to retrieve the camera pose, to perform a trifocal point transfer or obviously to compose the trifocal tensor itself.

We will now describe the linearization process itself. The Taylor expansion of a function $f(x,y)$ around the point $(0,0)$ is given by

$$ f(x,y) = f(0,0) + \frac{\partial f}{\partial x}(0,0) x + \frac{\partial f}{\partial y}(0,0) y + \cdots $$

where all terms of second or higher order have been omitted. The linearized result is the sum of the function at a known point (e.g. the rectified state) and the partial derivatives at this point multiplied by the deviation from the known state. In a similar way, we can express a rotation matrix as a function of three variables, i.e. the angles $\alpha_x$, $\alpha_y$, and $\alpha_z$. To clarify the linearization approach, we perform the calculation of the first order Taylor expansion of the rotation matrices explicitly. Subsequently, we will compute the projection matrices and the trifocal tensor.

The rotation matrix $R(\alpha_x, \alpha_y, \alpha_z)$ is a $3 \times 3$ matrix which is composed by trigonometric functions. However, we can assume that all angles $\alpha_x$, $\alpha_y$, and $\alpha_z$ are small as we are near the rectified state:

$$ R(\alpha_x, \alpha_y, \alpha_z) \approx \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} + \alpha_x \begin{bmatrix} 0 & 0 & 0 \\ 0 & -1 & 0 \\ 1 & 0 & 0 \end{bmatrix} + \alpha_y \begin{bmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ -1 & 0 & 0 \end{bmatrix} + \alpha_z \begin{bmatrix} 0 & -1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} $$

(2)

We will now use this result to calculate the linearized projection matrix.

$$ P = KR \left[ I \right] - C $$

(3)

We will first present the set of three intrinsic matrices $K, K'$, and $K''$ for the first, second, and third camera respectively. We assume that the focal lengths are similar and vary only by a small amount. The first camera is our reference camera. Assuming a centered origin, we get

$$ K = \begin{bmatrix} f & 0 & 0 \\ 0 & f & 0 \\ 0 & 0 & 1 \end{bmatrix} $$

(4)

In our case, the absolute focal length $f$ has little influence which allows us to approximate it [12] with $f = \sqrt{w^2 + h^2}$ where $w$ and $h$ are the image width and height respectively. We can now normalize $K$, i.e. the reference camera’s intrinsic matrix is the identity matrix $I$. The focal length of the second camera differs by a factor $1 + a'_y$ from the reference camera where $a'_y$ is small, i.e. second order terms can be neglected during linearization. The third camera is affected by a small horizontal shift of the principal point $p''_x$, a small offset $a''_r$ of its focal length, and a small deviation $a''_z$ of the aspect ratio from isotropy. These parameters are required to ensure an optimal horizontal alignment of the rectified cameras.

$$ K'(a'_y) = \begin{bmatrix} 1 + a'_y & 0 & 0 \\ 0 & (1 + a'_y) & 0 \\ (1 + a'_y)(1 + a''_r) & 0 & p''_x \end{bmatrix} $$

(5)

As the first camera is the reference camera, its center $C = (0,0,0)$ coincides with the origin. The second and third cameras lie on a common baseline, its centers are $C' = (c'_x, 0,0)$ and $C'' = (c''_x, 0,0)$ respectively. Please note that we have neglected translations of the cameras centers along the y-axis and the z-axis. The projection matrix for the first camera is $P = \left[ I \right]_0$. Given the camera centers $C'$ and $C''$ we get for the second and third cameras by inserting eqns. (5) and (2) into eq. (3) the projection matrices $P'$ and $P''$:

$$ P' = \begin{bmatrix} a'_x + 1 & -a'_z & a'_y \ 
-a'_z & a'_x + 1 & -c'_z(a'_y + 1) \ 
-a'_y & c'_z(a'_y + 1) & a'_x \end{bmatrix} $$

(6)

$$ P'' = \begin{bmatrix} a''_x & a''_y + 1 & -a'_z \ 
-a'_z & a''_x + 1 & -a''_y \ 
-c_z(a''_y + 1) & c_z(a''_y + 1) & a''_x \end{bmatrix} $$

Given the projection matrices, we can now perform a calculation of the trifocal tensor using eqns. (7,8) [6],

$$ T_{jk}^l = a'_l b'_k - a''_l b''_k $$

(7)

$$ P' = [a'_l] P'' = [b'_l] $$

(8)

where $a'_l$ and $b'_l$ represent the elements in the $l$th column and $j$th row of the projection matrix $P'$ and $P''$.
respectively. To increase the readability we substitute 
Δc_x = (c_x' − c_x''). We can now present the trifocal 
tensor in slices representation:

\[ T_1 = \begin{bmatrix} 1 + a'_1 + a''_1 + a''_2 \Delta c_x & a''_1 \Delta c_x & -a''_2 \Delta c_x \\ a_1 \Delta c_x + c'_2 & 0 & 0 \\ -a_1' \Delta c_x + c'_2 & 0 & 0 \\ -a_2' \Delta c_x + c'_2 & (1 + a'_1 + a''_1) c'_2 & a''_2 c'_2 \\ \end{bmatrix} \]

\[ T_2 = \begin{bmatrix} (a'_2 + p_x') c'_2 - a'_2 c'_2 & -a'_2 c'_2 & (1 + a'_1) c'_2 \\ a'_2 c'_2 & 0 & 0 \\ -a'_2 c'_2 & -a_2' c'_2 & a_2' c'_2 + a''_2 c'_2 \\ \end{bmatrix} \]

\[ T_3 = \begin{bmatrix} (a_2' + p_x') c'_2 - a'_2 c'_2 & -a'_2 c'_2 & (1 + a'_1) c'_2 \\ a'_2 c'_2 & 0 & 0 \\ -a_2' c'_2 & -a_2' c'_2 & a_2' c'_2 + a''_2 c'_2 \\ \end{bmatrix} \]

Given a set of triplet point matches, we do now aim to 
find the unknown variables numerically. Special 
attention is needed to estimate the components c''_1 and 
c''_2. In fact, all we need for a multi-baseline stereo 
approach is the ratio of the baselines between the three 
cameras. In consequence, we can normalize the 
baseline between the first and the second camera to 
β_{12} = 1 and concentrate on the estimation of the 
relative size of the baseline between first and third 
camera β_{13} = c''_1/c' and β_{13} = β_{12} + β_{23} . Given the 
trifocal tensor in slices representation and a set of 
triplet point matches, we can formulate the following 
set of nine equations:

\[ \begin{bmatrix} u' \\ v' \\ 1 \end{bmatrix} \begin{bmatrix} u'' \\ v'' \\ 1 \end{bmatrix} = 0_{3\times3} \]

We can group the coefficients to form a constraint 
matrix A forming a linear set of equations which 
allows us to find the vector x of geometric parameters:

\[ x = [a_y, a'_y, a''_y, a'_x, a''_x, a''_y, a'_y, a''_x, p'_x, p''_x] \]

The homographies shall transform the projection 
matrices such that they have the same orientation and 
intrinsic parameters as the reference camera P. Thus, we 
give the matrices

\[ P' = K'R'\left(\left[I - C_1'\right]\right); P'' = K''R''\left(\left[I - C_2'\right]\right) \]

we can formulate the following homographies H' and 
H'' using eqns. (4) and (5): 

\[ H' = K'(a'_x, a'_y, a'_z)^{-1}K'(a'_x, a'_y, a'_z)^{-1} \]

\[ H'' = K''(a''_x, a''_y, a''_z)^{-1}K''(a''_x, a''_y, a''_z)^{-1} \]

where R'(a'_x, a'_y, a'_z) and R'(a''_x, a''_y, a''_z) denote 
rotation matrices around the x-, y-, and z-axis respectively. 

Please note that the standard trigonometric functions 
are used and not a linearized version. After applying 
the homographies, the projection matrices have the 
desired simplified form corresponding to an ideal 
linear camera array:

\[ H'P' = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}; H''P'' = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix} \]

Please note that as the reference camera P remains 
unchanged during the rectification process, i.e. it can 
serve as reference camera for multiple camera triplets 
within the linear camera array. 

The precision of the geometric parameters can be 
improved by iterating the estimation process, i.e. the 
rectifying homographies are applied to the feature 
point triplets and a new Taylor expansion is 
performed. We iterate until no further improvement is 
achieved, i.e. if the back-projection error [6] for the 
resulting rectified projection matrices converges.

**Algorithm: Iterative Estimation of the Trifocal Tensor**

**Input:** 
- triplet point matches \([u, v, u', v', u'', v'']\) 
- Baseline ratios \([β_{12}, β_{13}, β_{23}]\) 

**Output:** 
- Final geometric parameter vector \(x_f\) 
- Projection Matrices \(P'\) and \(P''\) 
- Rectifying Homographies \(H'\) and \(H''\)

Initialize parameter vector \(x_f\) and baseline ratios \(x_f = \hat{b}; β_{12} = 1; β_{13} = β_{12} + β_{23}; m_1 = m_2; \) 

Try until back-projection error converges

\[ \text{do} \quad \text{Iterate until back-projection error converges} \]

\[ \text{do} \quad \text{Iterate until correction value } a_2 \text{ vanishes} \]

\[ \text{build new constraint matrix } A \text{ and solve eq.}(13) \]

\[ [A, b] = \text{get_constraint_matrix}(x_f, β_{13}, β_{23}, m_1) \]

\[ x_0 = \text{solve_linear_system}(A, b) \]

\[ β_{13} = x_0[\alpha_0]; \quad m_3 = \beta_{12} - β_{12} \]

\[ \text{until } \|x[α_0]\| < \varepsilon \]

\[ x_f = x_h + \text{Add parameter vector from last iteration} \]

\[ \text{Projection matrices according to eq. (14)} \]

\[ [P', P''] = \text{get_proj_matrices}(x_f) \]

\[ \text{Homographies according to eq. (15)} \]

\[ \{\text{Rectify homographies}\} \]

\[ \text{Apply homographies to feature points} \]

\[ m_3 = \text{get_rectified_points}(m_3, H', H'') \]

Calculate back-projection error after Rectification, eq. (16)

\[ \text{err}_{old} = \text{err}_{bp}; \quad \text{err}_{bp} = \text{back_proj_err}(m_3, H'P', H''P'') \]

\[ \text{until } |\text{err} - \text{err}_{old}| < \varepsilon \]
3. Results

In a first experiment, the proposed algorithm was tested using synthetic data. Varying noise amplitudes of the feature point positions and three different mechanical alignment qualities were simulated with subsequent analysis of the back-projection error and baseline ratios. For each angle $\alpha'_{\text{h}}, \alpha'_{\text{w}}, \alpha_{\text{h}}, \alpha_{\text{w}}$ a deviation of $1^\circ (5^\circ)$ was used to simulate low (high) rotation errors. The intrinsic parameters $(\alpha'_{\text{h}}, \alpha'_{\text{w}}, \alpha_{\text{h}}, \alpha_{\text{w}}, p_{\text{h}}, p_{\text{w}})$ were set to $f/100 (5 \times f/100)$ to simulate low (high) intrinsic errors. The camera centers were moved in $y$- and $z$-direction by 1% (5%) of the baseline $c_{\text{h}}$ to simulate low (high) translational errors. As shown in Table 1, the resulting back-projection errors are in the range of the noise amplitude and increase with higher mechanical alignment quality.

In a second experiment we used trinocular camera footage shot using three high quality HD cameras with a resolution of 1920×1080 pixels involving a narrow and a wide baseline to test the proposed algorithm. Fig. 2 (top) shows a rectified image triplet along with horizontal lines demonstrating that corresponding pixels lie on the same image scan line. Moreover, the disparity maps in inverse-of-a-distance representation [13] in Fig. 2 (bottom) demonstrate the quality of the horizontal alignment. In Fig. 3 the rectification accuracy is illustrated by analyzing the vertical and horizontal alignment before and after rectification of feature points estimated using [19].

Table 1. Back-projection error estimated for synthetic data with increasing noise level and alignment errors.

<table>
<thead>
<tr>
<th>Noise level $\sigma_{\text{n}}$ (in pixel)</th>
<th>Mechanical Alignment Quality</th>
<th>Back-Proj. Error $\text{err}_{\text{b/p}}$</th>
<th>Estimated Baseline Ratio $\beta_{13}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Good (low rot. &amp; intrinsic errors.)</td>
<td>Med. (low rot., intr. &amp; transl. err.)</td>
<td>Bad (high rot., intr. &amp; transl. err.)</td>
<td></td>
</tr>
<tr>
<td>0.0 0.000 / 5.000</td>
<td>0.311 / 5.000</td>
<td>1.554 / 4.989</td>
<td></td>
</tr>
<tr>
<td>0.1 0.086 / 5.000</td>
<td>0.331 / 5.000</td>
<td>1.558 / 4.989</td>
<td></td>
</tr>
<tr>
<td>0.2 0.172 / 5.000</td>
<td>0.371 / 4.999</td>
<td>1.571 / 4.989</td>
<td></td>
</tr>
<tr>
<td>0.5 0.431 / 4.999</td>
<td>0.548 / 4.998</td>
<td>1.646 / 4.988</td>
<td></td>
</tr>
<tr>
<td>1.0 0.864 / 4.996</td>
<td>0.924 / 4.994</td>
<td>1.846 / 4.987</td>
<td></td>
</tr>
<tr>
<td>2.0 1.733 / 4.983</td>
<td>1.765 / 4.981</td>
<td>2.439 / 4.970</td>
<td></td>
</tr>
</tbody>
</table>

4. Conclusion

We proposed a new method to rectify cameras in a linear array. The rectification parameters are estimated using a linearized trifocal tensor. The results show that the method based on feature point triplets from uncalibrated cameras is suitable to perform a vertical and horizontal alignment ensuring proportional horizontal disparities.

References

Figure 1. Evaluation of the rectification quality using synthetic data with a noise level of $\sigma_n = 0.5$ pixels and a medium mechanical alignment quality according to Table 1. The back-projection error ($err_{bp} = 0.548$) and the baseline ratios $\beta_{12} = 1.0, \beta_{23} = 3.998,$ and $\beta_{13} = 4.998$ were determined according to the proposed estimation algorithm. **Left:** The horizontal disparities of the unrectified synthetic data (original feature points marked in red) are not proportional which is the case instead for the rectified points (green), which lie mainly on the ideal blue line. **Center:** The horizontal disparities of the original feature points (red) have a considerable offset from the ideal position, e.g. proportionality. After rectification (green points), the offset has been minimized. **Right:** The original feature points (red) show vertical disparities. After rectification, the vertical disparities have been minimized.

Figure 2. Rectified multi-camera footage (top) along with normalized disparity maps (bottom). Horizontal reference lines illustrate that the vertical disparities vanished. The disparity maps were normalized using the baselines ratios $\beta_{12}, \beta_{23},$ and $\beta_{13}$ resulting in corresponding gray scale values for corresponding pixels despite non-equidistant baselines.

Figure 3. Evaluation of the rectification quality for the multi-camera footage from Fig. 2. The back-projection error ($err_{bp} = 0.298$) and the baseline ratios $\beta_{12} = 1.0, \beta_{23} = 4.004,$ and $\beta_{13} = 5.004$) were determined according to the proposed algorithm for the linearized estimation of the trifocal tensor. **Left:** The horizontal disparities of the original feature points (red) are not proportional which is the case for the rectified points (green), which lie mainly on the ideal blue line. **Center:** The horizontal disparities of the original feature points (red) have a considerable offset from the ideal position, e.g. proportionality. After rectification (green points), the offset has been minimized. **Right:** The original feature points (red) show vertical disparities. After rectification, the vertical disparities have been minimized.