1 Background

This report introduces some simplifications to the method by Fitzgibbon et al. [1] that allows for 3D model construction from turn-table sequences. It is assumed that the reader has previously read [1] in order to fully understand this report.

Fitzgibbon et al. presents a method for 3D model construction that utilizes the extra constraints imposed by turn-table sequences. Restricting the scenario to a turn-table sequence with a single camera with fixed settings produces these extra constraints:

C1. The internal parameters for the camera are the same for all images

C2. The motion of the camera can be described by a rotation around a single axis

It is shown that in the uncalibrated case the number of parameters to estimate is $m + 8$ where $m$ is the number of images.

We further simplify the problem by using extra constrains given from the fact that we know:

C3. The internal parameters of the camera, i.e the $K$ matrix

C4. That the angle between each pair of consecutive cameras is the same

Using these extra simplifications makes it possible to create a 3D model from realistic data without using Bundle Adjustment.

2 Method

The effects of the additional constraints C3-C4 are presented in this section. For full details regarding the original method see [1].
2.1 Original

The world coordinate system is chosen such that the rotation axis coincides with the z-axis. Using the notation in [1], the first camera may be written as

\[ P_0 = H[I|t] \]  \hspace{1cm} (1)

where \( H \) is a homography representing the camera internal parameters and a rotation about the camera center, and \( t = (t, 0, 0)^T \). The subsequent cameras are then given by post-multiplying \( P_0 \) with

\[
\begin{bmatrix}
R_z(\theta_i) & 0 \\
0 & 1
\end{bmatrix}
\]  \hspace{1cm} (2)

yielding the cameras

\[ P_{\theta_i} = H[R_z(\theta_i)|t] \]  \hspace{1cm} (3)

The paper [1] describes a method for estimating \( H \) and the different \( \theta_i \).

2.2 Simplified

Simplification C3 allows us to write (3) as

\[ P_{\theta_i} = KR_z(\theta_i)|t] \]  \hspace{1cm} (4)

Simplification C4 allows us to further write (4) as

\[ P_{\theta_i} = KR_z(i\theta)|t] \]  \hspace{1cm} (5)

where \( \theta \) is the known angle between each camera. These simplifications effectively reduce the number of unknowns from \( m + 8 \) to just the 3 parameters describing the rotation \( R \) since we assume the rotation \( \theta \) to be known.

2.3 Modifications to H-estimation

In order to estimate \( R \) will we still use the method for estimating \( H \) from [1] but with a few modifications.

Estimation of \( H \) basically comes down to estimating the image fixed entities \( l_s \), the projection of the rotation axis, and \( l_h \), the projection of the plane of rotation. Given these (estimated according to steps 1-4, Section 2.5 in [1]) \( H \) can be determined up to a 3-parameter solution

\[ H = [h_1, h_2, h_3] = [x_s, \mu x_s, \nu x_s + \omega d] \]  \hspace{1cm} (6)

where \( d = l_s \times (0, 0, 1)^T \) and \( \mu, \nu \) and \( \omega \) are the unknown parameters. It is possible to estimate \( \mu \) using the trifocal tensor and \( \nu \) and \( \omega \) using parallel scene lines.

Using known internal parameters, \( K \) allows us to simplify the estimation process. According to simplification C3, \( H \) can be written as

\[ H = KR \]  \hspace{1cm} (7)
leading to
\[ R = K^{-1}H. \] (8)

Since we cannot estimate \( H \) without resolving the 3-parameter ambiguity we cannot simply estimate \( R \) according to (8). However, we do have good estimates of the first two columns in \( H \) (where the second column is estimated only up to scale), using this we get the following
\[ [r_1, \mu r_2, 0] = K^{-1}[h_1, \mu h_2, 0]. \] (9)

Normalizing \( r_1 \) and \( r_2 \) allow us to estimate \( R \) as
\[ R = [\tilde{r}_1, \tilde{r}_2, \tilde{r}_1 \times \tilde{r}_2] \] (10)

since we know that we are looking for a rotation matrix \( R \in SO(3) \).

Due to measurement noise, \( r_1 \) and \( r_2 \) will not be perfectly orthogonal after (9). In order to remedy this we use a symmetric Gram-Schmidt orthogonalization procedure before computing (10). Ordinary Gram-Schmidt orthogonalization where \( \tilde{r}_1 \) is held still produces similar results.

One minor (but important) difference to the approach described in the paper is to average \( R \) from estimations between all in view-pairs, this gives considerably better result than just using one view-pair. It is also important to make sure that the averaged \( R \) is still a rotation matrix. This is easily achieved by taking the Singular Value Decomposition (SVD) of the sum of all rotations
\[ UDV^T = \sum R_i \] (11)

and then letting the averaged rotation be
\[ R = U \text{diag}(1, 1, \det(U) \det(V))V^T, \] (12)


3 Results

Two figures are shown below in order to visualize the performance of the method, see Fig. 1 and 2. The first figure shows the obtained 3D models when using the turn-table restrictions compared to creating the trajectory of cameras by the unconstrained ‘triangle-approach’ (introduced in the lectures). The second figure shows the estimated camera positions for both approaches. The results for the turn-table method were obtained on the given real trajectories (which contain real noise) while the result for the ‘triangle-approach’ were obtained on noise-free data with added Gaussian noise (with a standard deviation of 0.01 pixels). With the ‘triangle-approach’ on the real trajectories, it completely breaks down.
Figure 1: 3D model. Top: This method. Bottom: The unconstrained 'triangle-approach'.
Figure 2: Camera position. Top: This method. Bottom: The unconstrained 'triangle-approach'
4 Conclusions

The results obtained by using further restrictions on the turn-table sequence (known internal parameters and same rotation between each camera pair) but without the BA-step are not too bad, average reprojection error of 1.1 pixels. Feeding these results into a BA-step (not restricted to optimizing over the rotations only but with all parameters free) gives an average reprojection error of 0.01 pixels.

The results obtained without the BA-step are not perfect but it is at least possible to see that the reconstructed shape is not too strange, hence making this modified method suitable for a minor undergraduate project using the paper [1] combined with the modifications described in this paper.

References
