Turntable-Constrained Camera Pose Estimation

Norio Kosaka
LY Corporation
Tokyo, Japan
kosaka.norio@lycorp.co.jp

Shinichi Higashino
LY Corporation
Tokyo, Japan
shigashi@lycorp.co.jp

Shuji Yamaguchi
LY Corporation
Tokyo, Japan
shyamagu@lycorp.co.jp

Abstract

We study camera motion estimation for image sequences generated by single-axis rotation, such as turntable capture or orbiting cameras. In this setting, all views share a common rotation axis and differ primarily by a per-view rotation angle, forming a low-dimensional motion family. However, standard structure-from-motion (SfM) pipelines estimate unconstrained pairwise geometry and only afterwards attempt to fit an orbit, which can produce physically inconsistent trajectories when visual evidence is weak.

We show that under single-axis motion, the essential matrix has a structured form in which translation is induced by the rotation of a shared orbit vector. This reveals a low-dimensional family of feasible essential matrices that couple geometry across views. Based on this observation, we investigate estimators that progressively enforce the turntable constraint, including axis-projected rotations, structured essential matrix estimation, and global orbit refinement. Experiments on synthetic sequences and a Blender-based turntable dataset show that incorporating the single-axis prior improves rotation estimation accuracy and yields more physically consistent camera trajectories.

1. Introduction

Estimating camera motion from images is a central problem in 3D computer vision [1, 9, 13, 15]. With known intrinsics, the pose of a calibrated camera can be recovered from 2D–3D correspondences using PnP [14], while in the absence of 3D points structure-from-motion (SfM) jointly estimates camera poses and scene geometry from multi-view correspondences [6]. Modern SfM pipelines are highly successful across diverse scenes by combining pairwise epipolar geometry with global optimisation such as bundle adjustment. However, these pipelines typically treat each camera pose as an unconstrained six-degree-of-freedom variable. When the underlying motion follows a known physical process, ignoring this structure introduces unnecessary degrees of freedom and can lead to geometrically plausible but physically inconsistent reconstructions.

This paper studies one of the most common structured motion patterns in practical capture setups: *single-axis rotation*. Such motion arises whenever either the object or the camera rotates around a fixed axis. Typical examples include turntable-based object scanning, robotic inspection systems, and orbiting-camera capture used in photogrammetry and neural rendering pipelines. In these settings all camera centres lie on a circular orbit and each view differs primarily by a rotation angle about a shared axis. This strong geometric constraint reduces the dimensionality of the camera trajectory and induces distinctive projective structure in the multi-view geometry.

The geometry of single-axis motion has been studied since early turntable reconstruction work. Fitzgibbon *et al.* [3] showed that camera matrices in turntable sequences follow a restricted projective form and that rotation angles can be uniquely identified even under weak calibration assumptions. More recently, orbit-constrained optimisation [2] has been proposed as a way to recover physically meaningful trajectories when unconstrained SfM produces inconsistent camera motion. Despite these insights, most modern SfM pipelines do not explicitly incorporate the single-axis constraint during estimation. Instead, they estimate pairwise epipolar geometry using generic two-view solvers, recover camera poses in the full space of rigid motions, and only afterwards attempt to fit an orbit or axis to the reconstructed trajectory. When visual evidence is weak—e.g., due to low texture, repeated patterns, or partial rotational symmetry—this *estimate-then-project* strategy becomes brittle, as the high-dimensional pose parameterisation admits many solutions that explain noisy correspondences but violate the true motion. Figure 1 illustrates this behaviour on synthetic data that exactly follow a single-axis trajectory.

In this work we revisit camera motion estimation under single-axis rotation and study how the turntable constraint can be incorporated directly into geometric estimation. We show that under this motion model the essential matrix is not arbitrary: both the rotation and translation

are determined by a shared rotation axis and a single orbit vector, inducing a low-dimensional family of feasible essential matrices that links all image pairs across the sequence. Motivated by this structure, we investigate estimators that progressively enforce the turntable constraint during motion estimation. Starting from generic two-view pose estimation, we introduce axis-projected rotations, a structured essential-matrix formulation that couples rotation and translation, and a sequence-level orbit refinement stage. Together these steps move the estimation process from unconstrained pairwise geometry toward optimisation directly on the single-axis motion family.

Contributions. This paper revisits camera motion estimation under single-axis rotation and studies how the turntable constraint can be incorporated directly into geometric estimation. Our contributions are:

- We derive the structured form of the essential matrix induced by single-axis motion, revealing that both rotation and translation are coupled through a shared axis and orbit geometry.
- We propose a sequence of turntable-aware estimators that progressively enforce this constraint, including axis-projected rotations, structured essential-matrix estimation, and global orbit refinement.
- We provide a controlled evaluation protocol using synthetic and rendered turntable sequences, demonstrating that enforcing the single-axis prior improves rotation estimation accuracy and produces physically consistent trajectories compared to generic SfM baselines.

2. Related Work

We briefly review prior work related to turntable and single-axis reconstruction, orbit-constrained optimisation, and generic structure-from-motion pipelines. These works highlight both the geometric structure induced by circular motion and the limitations of existing SfM pipelines that do not explicitly enforce this constraint during estimation.

2.1. Single-Axis and Turntable Reconstruction

Single-axis (turntable) motion has long been exploited to simplify multiview reconstruction and pose estimation. Fitzgibbon *et al*. derive the projective geometry induced by uncalibrated single-axis rotation, showing that camera matrices share a restricted form, rotation angles are uniquely recoverable, and reconstruction is determined up to a two-parameter projective ambiguity along the rotation axis [3]. Closely related analyses study orbital camera motion and the recovery of structure and calibration when cameras follow circular trajectories [12].

Trajectory-based formulations further exploit the fact that rotational motion induces conic image trajectories. Sawhney *et al*. group tracked points into coherent rotational trajectories and recover 3D structure and motion from these trajectories under perspective projection [18], while Szeliski studies recovering shape from rotation as a constrained motion family [20]. Under known calibration, single-axis motion implies that each 3D point traces a circle whose image is a conic. Fremont and Chellali exploit this structure to recover 3D point coordinates from conic observations followed by reprojection-error refinement [5], and show that camera calibration can be obtained from concentric circles in turntable capture [4]. More generally, Jiang *et al*. demonstrate that conic fitting over long point tracks can recover the geometry of single-axis motion even when rotation angles are unknown [10, 11].

These approaches leverage strong geometric regularities induced by circular motion. However, they typically rely on accurate long-term feature tracking or precise conic estimation, which can be difficult in practical capture scenarios with noisy correspondences or missing observations.

2.2. Orbit-Constrained Optimisation

More recent work incorporates orbit constraints directly into optimisation. Elms *et al*. propose an orbit-constrained bundle adjustment framework that jointly estimates camera centres, orbit parameters, and 3D structure within a factor-graph formulation [2]. In their model the camera trajectory follows a circular orbit parameterised by a single rotational rate f, so that the camera angle evolves linearly with time. They show that constraining optimisation to this motion family yields physically meaningful trajectories even when unconstrained SfM achieves lower reprojection error.

However, the assumption of a constant angular rate restricts the motion model and may not accurately reflect real capture setups where rotation speed varies. Furthermore, orbit parameters are typically initialised from a generic SfM reconstruction, so the motion constraint is imposed only after unconstrained pose estimation.

2.3. Generic SfM and Motion Priors

Modern structure-from-motion systems estimate camera motion by first recovering pairwise epipolar geometry and then performing global optimisation such as bundle adjustment or rotation averaging [7, 16]. Widely used pipelines such as COLMAP implement robust incremental SfM with global refinement [19].

These pipelines treat each camera pose as an independent variable and estimate pairwise essential matrices without explicitly enforcing global motion constraints. While highly effective for general scenes, they do not exploit the strong geometric structure of single-axis motion. As a result, reconstructed trajectories may deviate from the physically plausible orbit and motion constraints must be enforced only post-hoc through trajectory fitting or con-

strained optimisation.

This issue becomes particularly pronounced in turntable capture scenarios. Such setups are often performed in controlled studio environments with static or featureless backgrounds, and objects may exhibit partial or full rotational symmetry. In these cases image correspondences become ambiguous and provide weak constraints on epipolar geometry. Consequently, unconstrained pose estimation can become ill-conditioned even though the underlying motion follows a simple low-dimensional model.

These observations motivate incorporating the single-axis motion constraint directly into the estimation of epipolar geometry rather than enforcing it only after generic pose recovery.

3. Problem Statement

Single-axis capture is common in turntable scanning and orbiting-camera setups, where either the object rotates on a turntable or the camera moves around a stationary object. As discussed in Section 2, such motion follows a highly structured geometric model that significantly reduces the degrees of freedom of the multi-view reconstruction problem. In this section we formalise the single-axis motion model and derive the structured form of the essential matrix induced by this motion.

3.1. Single-Axis Motion Geometry

We observe m perspective views of a rigid scene undergoing a *single-axis rotation*, corresponding either to an object rotating on a turntable or an orbiting camera [2, 3, 7]. Let $x_{ik} \in \mathbb{P}^2$ denote the image of 3D point $X_k \in \mathbb{P}^3$ in view i. Camera intrinsics K may be known, partially known, or unknown.

Single-axis motion model. All camera motions are rotations about a fixed 3D axis with no translation along that axis. Let $a \in \mathbb{S}^2$ denote the rotation axis direction, and let $R(a, \theta_i) \in \mathrm{SO}(3)$ be the rotation of angle θ_i about a. The camera centres lie on a circle orthogonal to a and admit the decomposition

$$C_i = C_{\parallel} + R(a, \theta_i)\, C_{\perp}, \qquad C_{\perp} \perp a, \tag{1}$$

where $C_{\parallel}$ is the component along the axis and $\|C_{\perp}\|$ is the (unknown) orbit radius.

The camera orientations are consistent with the same axial rotation,

$$R_i = R_0\, R(a, \theta_i), \tag{2}$$

where R_0 encodes a fixed but unknown reference orientation.

Relative motion. Between two views i and j, the relative rotation and translation satisfy

$$R_{ij} = R(a, \theta_j - \theta_i), \qquad t_{ij} = (I - R_{ij})\, C_{\perp}. \tag{3}$$

Thus translation is not an independent variable but is induced by the rotation of the orbit vector $C_{\perp}$. Unlike prior work [2], we do not assume a fixed rotation frequency. The model assumes negligible translation along the rotation axis, which is a reasonable approximation for turntable setups, though small deviations may occur in practice.

Each view is represented by a perspective camera

$$P_i \sim K\,[\,R_i \mid -R_i C_i\,], \tag{4}$$

and image measurements satisfy the perspective projection relation

$$x_{ik} \sim \pi\left(P_i \tilde{X}_k\right), \tag{5}$$

where $\pi(\cdot)$ denotes projection from homogeneous image coordinates to $\mathbb{P}^2$.

Unknowns. The unknown parameters are:
- per-view rotation angles $\{\theta_i\}_{i=1}^{m}$ (not assumed uniform);
- the rotation axis a and orbit geometry $(C_{\parallel}, C_{\perp})$;
- camera intrinsics K (depending on the calibration setting);
- 3D structure $\{X_k\}$.

Identifiability. Under single-axis motion the multi-view geometry admits a reduced parameterisation. Instead of $11m$ camera parameters in generic projective SfM, the camera matrices depend on $m+8$ degrees of freedom. The rotation angles θ_i are uniquely identifiable from image measurements, while the reconstruction remains determined only up to a two-parameter projective ambiguity corresponding to a one-dimensional projective transformation along the rotation axis.

3.2. Structured Essential Matrix under Single-Axis Motion

For calibrated cameras (*i.e.* the intrinsic matrix K is known), the relative motion between two views (i, j) is described by the essential matrix

$$E_{ij} = [t_{ij}]_{\times} R_{ij}, \tag{6}$$

where $R_{ij} \in \mathrm{SO}(3)$ is the relative rotation and $t_{ij} \in \mathbb{R}^3$ is the translation direction between the camera centres.

Under the single-axis motion model introduced in Section 3.1, the relative rotation between two views satisfies

$$R_{ij} = R(a, \theta_j - \theta_i). \tag{7}$$

The corresponding translation is induced by the rotation of the orbit vector $C_\perp$,

$$t_{ij} = (I - R_{ij})\, C_\perp, \qquad C_\perp \perp a. \tag{8}$$

Substituting these expressions into the essential matrix definition yields

$$E_{ij} = \left[(I - R(a, \theta_j - \theta_i))\, C_\perp\right]_\times R(a, \theta_j - \theta_i). \tag{9}$$

Equation (9) shows that the essential matrix is not arbitrary: both the rotation and translation are determined by the shared axis a, the orbit vector $C_\perp$, and the relative rotation angle $\Delta\theta_{ij} = \theta_j - \theta_i$. Consequently, the set of feasible essential matrices forms a low-dimensional structured family induced by the turntable motion.

This coupling has two important implications. First, translation is no longer an independent variable but is completely determined by the rotation and orbit geometry. Second, all image pairs share the same global parameters $(a, C_\perp)$, linking their epipolar geometries across the sequence.

In the following sections we exploit this structure to design estimators that progressively enforce the single-axis prior during motion estimation.

3.3. Empirical Demonstration of the Issue

To illustrate the practical difficulty of estimating motion in turntable scenarios, we apply a standard SfM pipeline to a controlled orbiting-camera setup.

We render a sequence of $m = 100$ images of a textured 3D dog model using Blender, where the camera follows a circular trajectory around the object with a fixed rotation axis and non-uniform angular increments.

We run COLMAP using the default incremental SfM pipeline with calibrated intrinsics. Despite the fact that the ground-truth motion satisfies the single-axis model exactly, the recovered camera poses deviate significantly from the true circular trajectory. Figure 1 shows the resulting reconstruction, where the estimated camera centres fail to lie on a common orbit.

This failure arises because generic SfM pipelines estimate unconstrained pairwise epipolar geometry and treat camera poses as independent variables, ignoring the low-dimensional structure of single-axis motion. Consequently, the recovered trajectory may deviate from the physically plausible orbit even when the underlying motion strictly follows the turntable model.

This example highlights the need for estimation methods that explicitly incorporate the single-axis motion constraint during pose estimation. Details of the experiments can be found in SM. 1.5.

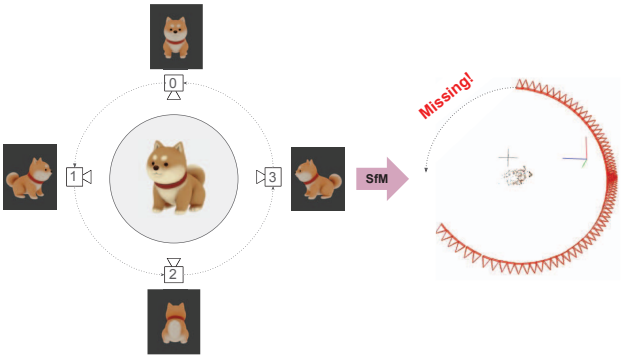

Figure 1. Failure of COLMAP on an orbiting-camera sequence with single-axis motion.

4. Approach

We study relative pose estimation for image sequences generated by single-axis rotation. Although this motion restricts camera poses to a low-dimensional orbit family (Sec. 3.1), most SfM pipelines estimate pairwise geometry using generic two-view solvers and enforce the orbit constraint only post-hoc. We therefore investigate how progressively enforcing the turntable constraint affects pose estimation.

We therefore consider two classes of estimators: (i) generic baselines that do not incorporate the motion prior during estimation, and (ii) turntable-aware variants that explicitly enforce the shared-axis structure of the motion.

4.1. Baselines: Generic Pose Estimation

We first consider standard geometric estimators that treat each camera motion as independent. These methods serve as reference points for evaluating the benefit of turntable-aware constraints.

A: Adjacent Essential-Matrix Decomposition. For each adjacent image pair $(i, i+1)$ we estimate the fundamental matrix $F_{i,i+1}$ from point correspondences using the normalized eight-point algorithm [8]. With known intrinsics K, the essential matrix is obtained as

$$E_{i,i+1} = K^\top F_{i,i+1} K. \tag{10}$$

We enforce the essential constraints by projecting $E_{i,i+1}$ onto the space of rank-two matrices with singular values $(1, 1, 0)$. By definition (6), decomposition of $E_{i,i+1}$ yields the relative rotation $R_{i,i+1} \in \mathrm{SO}(3)$ and translation direction. The incremental rotation is expressed in axis–angle form

$$R_{i,i+1} = \exp(\Delta\theta_i [\mathbf{u}_i]_\times). \tag{11}$$

A global axis estimate can then be obtained by averaging $\{\mathbf{u}_i\}$ across pairs and integrating the incremental angles.

This estimator corresponds to classical calibrated two-view geometry and does not exploit the single-axis motion prior.

recoverPose: OpenCV Relative Pose Baseline. We additionally use the OpenCV `recoverPose` pipeline. For each adjacent pair, the essential matrix is estimated with RANSAC, and the relative pose is recovered by selecting the cheirality-consistent decomposition with the largest support. Like Method A, this approach operates on unconstrained two-view geometry and does not enforce the shared-axis structure of turntable motion.

B: Windowed Rotation Averaging. To improve robustness, we estimate relative rotations R_{ij} for multiple pairs (i, j) within a temporal window and recover absolute rotations $\{R_i\}$ via rotation averaging:

$$\min_{\{R_i \in \mathrm{SO}(3)\}} \sum_{(i,j)\in\mathcal{E}} \|R_j - R_i R_{ij}\|_F^2. \tag{12}$$

We follow the chordal formulation with projection onto $\mathrm{SO}(3)$ as in [7]. Incremental rotations are recovered as $R_{i,i+1} = R_i^\top R_{i+1}$. Although this approach improves global consistency, it still treats each rotation as an unconstrained element of $\mathrm{SO}(3)$.

C: Homography-Based Rotation Approximation. As a degenerate baseline, we estimate a homography $H_{i,i+1}$ between adjacent views and approximate the rotation via

$$R_{i,i+1} \approx \Pi_{\mathrm{SO}(3)}(K^{-1} H_{i,i+1} K), \tag{13}$$

where $\Pi_{\mathrm{SO}(3)}(\cdot)$ denotes projection onto $\mathrm{SO}(3)$. This approximation is valid only under planar scenes or pure rotational motion and serves primarily as a weak geometric reference.

4.2. Propositions: Turntable-Aware Variants

We now introduce estimators that explicitly exploit the single-axis motion structure. These methods progressively enforce the constraint that all camera motions correspond to rotations about a shared axis.

A+: Projection onto the Single-Axis Subgroup. Given relative rotations $\hat{R}_{i,i+1}$ estimated by Method A and a rotation-axis estimate $a \in \mathbb{S}^2$, we project each rotation onto the one-parameter subgroup of rotations about a:

$$R^+_{i,i+1} = R_a(\Delta\theta_i), \tag{14}$$

where

$$\Delta\theta_i = \arg\min_\theta \|\hat{R}_{i,i+1} - R_a(\theta)\|_F^2. \tag{15}$$

Let (u, v, a) form an orthonormal basis with $u, v \perp a$. The optimal angle is then

$$\Delta\theta_i = \mathrm{atan2}(v^\top \hat{R}_{i,i+1} u,\ u^\top \hat{R}_{i,i+1} u). \tag{16}$$

This step enforces that all incremental rotations lie on the same axis, but does not yet couple the translation components.

A++: Turntable-Coupled Essential Matrix. Under the single-axis model, relative translation is not independent but determined by the orbit vector ρ orthogonal to the axis:

$$t_{i,i+1} = (I - R_a(\Delta\theta_i))\rho, \qquad \rho \perp a. \tag{17}$$

Substituting into the essential matrix definition yields

$$E_{i,i+1} = [(I - R_a(\Delta\theta_i))\rho]_\times R_a(\Delta\theta_i). \tag{18}$$

This structured essential matrix couples rotation and translation across all pairs through the shared orbit geometry.

A+++: Global Orbit Refinement. Finally, we refine the entire sequence by jointly optimising the orbit parameters, camera angles, and optionally the 3D structure via a turntable-constrained bundle adjustment:

$$\min_{\Theta, \{X_k\}} \sum_{i,k} \rho(\|x_{ik} - \pi(P_i(\Theta) X_k)\|^2). \tag{19}$$

This stage enforces global consistency of the orbit trajectory and corresponds to optimisation directly on the single-axis motion family.

5. Experiments

Our experiments are designed to evaluate the effect of explicitly enforcing the single-axis motion prior in turntable sequences. In particular, we study: (i) whether turntable-aware estimators improve rotation accuracy compared to generic two-view and SfM baselines; (ii) how the methods behave under controlled conditions where the ground-truth axis, trajectory, and structure are known; and (iii) whether the same trends persist on realistic image data where correspondences are obtained from feature matching rather than direct projection. Additional ablations studying sequence length, correspondence support, initialization effects, and baseline regimes are provided in SM. 2.

Metrics We evaluate rotation accuracy and trajectory consistency using standard geometric metrics. For pairwise motion estimation, we measure **relative rotation error** and

report **AUC@5°, AUC@10°, and AUC@20°**. When absolute camera orientations are available, we additionally report **mean absolute rotation error**. For methods that produce a full camera trajectory, we evaluate trajectory quality using **translation-direction relative pose error (RPE)**, **trajectory coverage error**, and **normalized loop-closure error**. Detailed metric definitions and implementation details are provided in SM. 1.2.

5.1. Synthetic Images

Data We evaluate the methods on a controlled synthetic orbiting-camera dataset generated using a parametric turntable model. A calibrated pinhole camera follows a circular trajectory around the scene with a fixed rotation axis, while 3D points are randomly sampled in front of the camera and projected into multiple views with additive Gaussian image noise. This setup provides ground-truth rotations, trajectory geometry, and full point correspondences, enabling precise evaluation of rotation and trajectory accuracy. Detailed dataset parameters and generation procedures are provided in SM. 1.1.

Results Figure 2 summarises rotation accuracy on the synthetic orbiting-camera dataset as the image noise level increases. We report AUC@10, AUC@20, and mean absolute rotation error for all methods.

At low noise levels, several calibrated two-view baselines already achieve very accurate rotation estimates. As the noise level increases, all methods degrade gradually, but the relative behaviour between estimators becomes more pronounced.

The homography baseline (C) performs substantially worse than the other methods across all noise levels. This behaviour is expected because the motion involves general 3D rotation with depth variation rather than planar geometry or pure in-place rotation.

The turntable-aware variants remain competitive with the strongest baselines throughout the noise range. In particular, the axis-projected estimator (A+) and the more constrained variants (A++ and A+++) closely track the performance of standard two-view geometry while enforcing the shared-axis motion model. These results indicate that incorporating the single-axis motion structure maintains strong rotation accuracy while preserving physically consistent trajectories.

Runtime Analysis We report runtime on the synthetic benchmark under the default configuration. All experiments were executed on a CPU (Intel Core Ultra 7 165U, 16 GB RAM) without GPU acceleration.

Table 1 summarises average runtime per sequence. Rotation averaging (B) is the fastest method (30–70 ms), while two-view and RANSAC-based methods such as `recoverPose` and the A-family range from 100 ms to 3000 ms depending on noise.

Method	Runtime (ms)
B	30–70
recoverPose	100–3000
A	200–3000
A+	200–3000
A++	450–3400
A+++	300–3200
8pt	900–1300

Table 1. Typical runtime ranges per sequence on the synthetic benchmark.

Among the proposed methods, A+ and A+++ have similar runtime to A, while A++ is the most expensive (450–3400 ms) due to the coupled optimisation. Runtime increases with image noise, as noisier correspondences lead to more RANSAC iterations.

Overall, the additional constraints introduce moderate overhead, and runtime is dominated by correspondence estimation and robust model fitting.

5.2. Rendered Turntable Experiments

To evaluate the proposed turntable-aware estimators on realistic image data, we construct a controlled orbiting-camera dataset using Blender. The goal is to reproduce the visual conditions commonly encountered in studio-style object capture while retaining full control over the motion model and rendering environment.

Rendered Turntable Dataset. We evaluate the proposed methods on a rendered turntable dataset constructed in Blender. Five publicly available 3D object models—*Shiba*, *Car*, *Aeroplane*, *Laptop*, and *Coffee Maker*—are placed at the center of a virtual turntable and rendered while undergoing a full single-axis rotation. For each object we generate a sequence of 100 images uniformly sampling a full revolution with fixed camera intrinsics and a simple studio-style lighting setup that mimics controlled object-capture conditions.

The objects are chosen to cover diverse geometric structures and partial symmetries, exposing common failure modes of generic SfM pipelines under single-axis motion while maintaining a controlled and reproducible evaluation setting. Figure 3 shows example objects from the dataset. Additional details on model sources, rendering configuration, and lighting setup are provided in SM. 1.3.

Experimental Pipeline Feature correspondences are extracted using ORB features (up to 3000 keypoints per image) and matched with a brute-force Hamming matcher

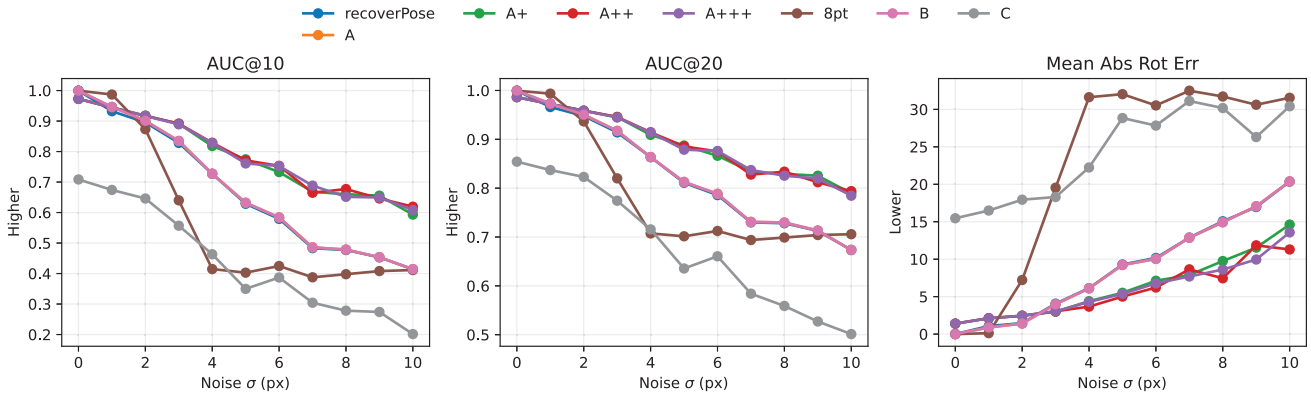

Figure 2. Rotation estimation performance on the synthetic turntable dataset as image noise increases. We report AUC@10°, AUC@20°, and mean absolute rotation error for all estimators. Turntable-aware variants maintain competitive accuracy while enforcing the shared-axis motion constraint.

Figure 3. Example objects used in the rendered turntable dataset.

with cross-checking. The essential matrix is estimated using RANSAC and decomposed into relative pose via a cheirality check. Additional implementation details are provided in SM. 1.4. The baseline and turntable-aware estimators are then applied to recover incremental rotations and the shared rotation axis, from which the global camera trajectory is obtained.

The rendering setup introduces realistic image measurements with shading and texture variations, and correspondences are obtained through standard feature matching pipelines rather than direct access to ground-truth projections. All evaluation metrics and estimator configurations remain consistent with those used throughout the paper.

Results Table 2 reports rotation estimation performance averaged across the five rendered turntable sequences; the full table is provided in SM. 2.7. Generic two-view baselines achieve moderate accuracy but struggle at strict angular thresholds. For example, `recoverPose` obtains only 0.053 AUC@5°, while COLMAP slightly improves this to 0.110, indicating that many estimated rotations are approximately correct but lack high-precision consistency.

Rotation averaging (Method B) improves robustness relative to direct two-view estimation, achieving 0.172 AUC@5° and 0.582 AUC@20°, which highlights the benefit of enforcing global rotational consistency across multiple pairs. However, because these methods estimate rotations in the full $\mathrm{SO}(3)$ space, they remain unconstrained with respect to the underlying single-axis motion.

The homography baseline (Method C) performs substantially worse than other baselines, producing both low AUC scores and the largest mean rotation error. This behaviour is expected since the motion involves general 3D rotation with depth variation, which cannot be accurately modeled by a planar homography.

In contrast, the proposed turntable-aware estimators progressively improve rotation accuracy as more structure is incorporated. Projecting pairwise rotations onto a shared axis (A+) increases AUC@5° from 0.172 to 0.198 and improves performance across all thresholds. Incorporating the structured essential-matrix formulation (A++) further stabilizes estimation and reduces the mean rotation error relative to the simple axis-projection approach.

The largest improvement is obtained when the entire sequence is refined under the turntable model (A+++), which achieves the best AUC scores across all thresholds (0.335 AUC@5°, 0.538 AUC@10°, 0.734 AUC@20°). Although the mean absolute rotation error does not consistently decrease, the higher AUC values indicate a more concentrated error distribution. This suggests that the proposed constraints reduce moderate errors while occasional larger errors persist, indicating some sensitivity to outliers. Overall, these results demonstrate that explicitly enforcing the single-axis motion prior improves rotation estimation reliability in controlled turntable-style image sequences.

Table 2. Rotation estimation results averaged over sequences. Higher is better for Pair Coverage and AUC; lower is better for Mean Absolute Rotation Error. Bold indicates best overall; underlined indicates best baseline.

Method	Pair Cov. ↑	AUC@5° ↑	AUC@10° ↑	AUC@20° ↑	Mean Abs. Rot. Err. ↓
8pt	**0.985**	0.158	<u>0.367</u>	0.575	93.55
recoverPose	0.864	0.053	0.164	0.452	73.79
COLMAP	0.858	0.110	0.245	0.503	**73.39**
B	**0.985**	<u>0.172</u>	0.345	<u>0.582</u>	81.50
C	**0.985**	0.085	0.205	0.435	97.92
A	**0.985**	<u>0.172</u>	0.345	<u>0.582</u>	81.50
A+	**0.985**	0.198	0.383	0.619	76.93
A++	**0.985**	0.183	0.361	0.599	76.22
A+++	**0.985**	**0.335**	**0.538**	**0.734**	85.74

Limitations. All experiments are conducted on synthetic or rendered turntable sequences, enabling controlled evaluation but not fully capturing real-world deviations such as calibration noise, imperfect axes, or small axial translations. Validation on real capture systems is left for future work.

6. Conclusion

We studied camera motion estimation under single-axis (turntable) rotation and showed that this motion induces a structured family of essential matrices in which rotation and translation are coupled through a shared axis and orbit geometry. Based on this structure, we investigated estimators that progressively enforce the turntable constraint, from axis-projected rotations to structured essential-matrix estimation and global orbit refinement. Experiments on synthetic and rendered turntable sequences demonstrate that incorporating the single-axis prior improves rotation accuracy and produces more physically consistent camera trajectories than generic SfM baselines. Overall, this work highlights the benefits of estimating directly on the turntable motion family and provides a foundation for future minimal solvers and robust turntable-constrained SfM methods.

References

[1] Eric Brachmann, Alexander Krull, Sebastian Nowozin, Jamie Shotton, Frank Michel, Stefan Gumhold, and Carsten Rother. Dsac-differentiable ransac for camera localization. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 6684–6692, 2017. 1

[2] Ethan Elms, Yasir Latif, Tae Ha Park, and Tat-Jun Chin. Event-based structure-from-orbit. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 19541–19550, 2024. arXiv:2405.06216. 1, 2, 3

[3] Andrew W. Fitzgibbon, Geoff Cross, and Andrew Zisserman. Automatic 3D model construction for turn-table sequences. In *3D Structure from Multiple Images of Large-Scale Environments*, pages 155–170. Springer, 1998. 1, 2, 3

[4] Vincent Frémont and Ryad Chellali. Direct camera calibration using two concentric circles from a single view. In *Proceedings of the 12th International Conference on Artificial Reality and Telexistence (ICAT)*, 2002. December 4–6, Tokyo, Japan. 2

[5] Vincent Frémont and Ryad Chellali. Turntable-based 3D object reconstruction. In *Proceedings of the 2004 IEEE Conference on Cybernetics and Intelligent Systems (CIS)*, pages 1277–1282, 2004. 2

[6] Richard Hartley and Andrew Zisserman. *Multiple view geometry in computer vision*. Cambridge university press, 2003. 1

[7] Richard Hartley, Jochen Trumpf, Yuchao Dai, and Hongdong Li. Rotation averaging. *International Journal of Computer Vision*, 103(3):267–305, 2013. 2, 3, 5

[8] Richard I Hartley. In defense of the eight-point algorithm. *IEEE Transactions on pattern analysis and machine intelligence*, 19(6):580–593, 1997. 4

[9] Joao F Henriques and Andrea Vedaldi. Mapnet: An allocentric spatial memory for mapping environments. In *proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 8476–8484, 2018. 1

[10] Guang Jiang, Hung-Tat Tsui, Long Quan, and Andrew Zisserman. Single axis geometry by fitting conics. In *Computer Vision – ECCV 2002*, pages 537–550. Springer, 2002. 2

[11] Guang Jiang, Hung-Tat Tsui, Long Quan, and Andrew Zisserman. Geometry of single axis motions using conic fitting. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 25(10):1343–1348, 2003. 2

[12] Sing Bing Kang. Quasi-euclidean recovery from unknown but complete orbital motion. Technical Report CRL 97/10, Digital Equipment Corporation, Cambridge Research Laboratory, 1997. 2

[13] Alex Kendall, Matthew Grimes, and Roberto Cipolla. Posenet: A convolutional network for real-time 6-dof camera relocalization. In *Proceedings of the IEEE international conference on computer vision*, pages 2938–2946, 2015. 1

[14] Vincent Lepetit, Francesc Moreno-Noguer, and Pascal Fua. Ep n p: An accurate o (n) solution to the p n p problem. *International journal of computer vision*, 81(2):155–166, 2009. 1

[15] David Nistér. An efficient solution to the five-point relative pose problem. *IEEE transactions on pattern analysis and machine intelligence*, 26(6):756–770, 2004. 1

[16] Onur Özyeşil, Vladislav Voroninski, Ronen Basri, and Amit Singer. A survey of structure from motion. *Acta Numerica*, 26:305–364, 2017. 2

[17] Maxime Raafat. Blendernerf, 2024. 2, 5

[18] Harpreet S. Sawhney, John Oliensis, and Allen R. Hanson. Image description and 3-D reconstruction from image trajectories of rotational motion. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 15(9):885–898, 1993. 2

[19] Johannes L. Schönberger and Jan-Michael Frahm. Structure-from-motion revisited. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 4104–4113, 2016. 2

[20] Richard Szeliski. Shape from rotation. In *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 625–631, 1991. 2