

Articulated and Restricted Motion Subspaces and Their Signatures

Bastien Jacquet¹,Roland Angst^{12*},Marc Pollefeys¹ETH Zürich¹, SwitzerlandStanford University², CA, USA

Abstract

Articulated objects represent an important class of objects in our everyday environment. Automatic detection of the type of articulated or otherwise restricted motion and extraction of the corresponding motion parameters are therefore of high value, e.g. in order to augment an otherwise static 3D reconstruction with dynamic semantics, such as rotation axes and allowable translation directions for certain rigid parts or objects. Hence, in this paper, a novel theory to analyse relative transformations between two motion-restricted parts will be presented. The analysis is based on linear subspaces spanned by relative transformations. Moreover, a signature for relative transformations will be introduced which uniquely specifies the type of restricted motion encoded in these relative transformations. This theoretic framework enables the derivation of novel algebraic constraints, such as low-rank constraints for subsequent rotations around two fixed axes for example. Lastly, given the type of restricted motion as predicted by the signature, the paper shows how to extract all the motion parameters with matrix manipulations from linear algebra. Our theory is verified on several real data sets, such as a rotating blackboard or a wheel rolling on the floor amongst others.

1. Introduction

Rigid structure-from-motion (SfM) has developed to a mature state in recent years and can provide convincing and highly accurate 3D reconstructions. Furthermore, 3D point clouds of indoor scenes can nowadays be quickly captured with cheap depth cameras. Unfortunately, even though the majority of a scene is generally indeed rigid, the aforementioned techniques will reconstruct all the objects in a scene in a static state and rigidly attached to the background. Separate objects or object parts and their dynamic relation will therefore be lost in this process. However, the observation of articulated or otherwise restricted motions between objects can provide valuable information about the dynamic relationship between these objects and parts and ultimately also about semantic classes of objects. In this paper, we

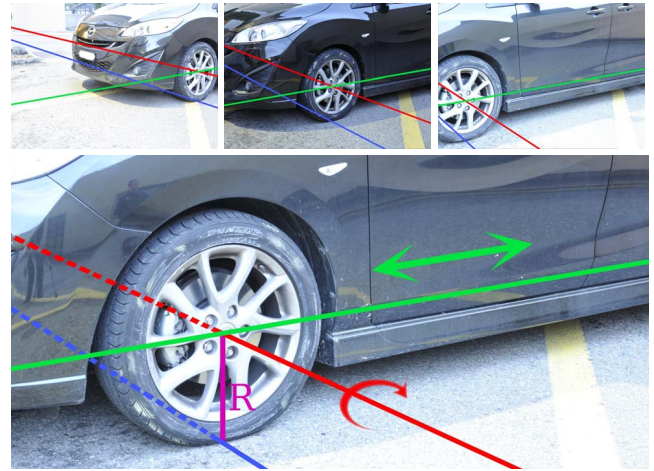


Figure 1: Application example : The framework in this paper enables the automatic extraction of the rotation axis (in red) and the translation direction (in green) from rigid transformations of a wheel rolling on the street. Moreover, as a side results, the wheel radius (in pink) and its contact line (in blue) with the ground can be inferred, as well.

therefore focus on such motions and the primary goal is to automatically detect the type of articulated or restricted motion class between two parts or objects and extract all relevant parameters of these restricted motions. We refer to Fig. 1 for an example application where the motion parameters of a wheel rolling on the street have been extracted automatically. In order to achieve this, we propose to analyse the relative rigid transformations between two parts. Note that the computation of these relative transformations is not the main objective of this work. These transformations can for example be computed with existing methods such as an iterative-closest-point algorithm applied to point clouds measured by depth cameras, or a classical SfM approach working on pure image data as we did in our experiments. The paper will explain that the relative transformations can be arranged in a single *motion matrix* which encodes all the information for determining which type of restricted motion has been observed and for computing all the relevant parameters of that motion.

*This work was done while this author was employed by ETH Zürich

Contribution: The main contribution of the present paper is the introduction of a so-called *signature* for relative transformations between two parts. This signature is a function of the *motion matrix* which is entirely determined by considering observed transformations as data samples from a linear subspace. The signature uniquely describes the type of restricted motion, such as a planar motion where an object translates on a plane and rotates around rotation axes which are orthogonal to this plane. Different signatures basically represent an extensive catalogue of restricted motions in the sense that they enumerate various types of motions together with algebraic constraints which have to be met by a certain type of restriction. Besides motivating and deriving properties of this signature, we will also show how all the relevant parameters of a restricted motion can be extracted by simply solving linear systems of equations. In addition to subsuming well-known cases of articulated motions (such as rotations around a fixed point), our derivations will lead to an unified framework which also covers novel types of restricted motions, *e.g.* articulated motions around two non-intersecting fixed rotation axes or around a translating rotation axis with fixed orientation can be treated in the very same way. As we will see in the experiments, these novel types of restricted motions are practically highly relevant. Note that in cases where not every individual part can be directly observed (*e.g.* due to lack of trackable features or occlusion), the resulting observable motions can become a complex chain of individual transformations. Being a unique feature of our analysis, our method can in particular detect a sequence of two subsequent rotations and untangle these so that the motion of a potential intermediate part can be hypothesized accurately.

2. Related Work

We will mainly discuss related work about articulated motions and subspace representations for SfM, since our method relies on those techniques. We refer to [10, 1] and references therein for general rigid SfM.

The analysis of articulated motions has been a topic of active research since many years. The importance of articulated skeletons for approximating the motion of a human body has often motivated research about articulated motions in computer vision or graphics [14, 13, 6, 21, 16, 15], whereas in robotics, teaching robots how to interact with man-made environments has been a driving force [12, 17]. More specifically, O’Brien *et al.* [14] measured relative transformations between articulated parts with a magnetic motion capture system. Assuming two parts rotate around a common fixed joint, this joint represents a fixed point under the relative transformations and can be computed with linear methods. This observation is actually the major reoccurring concept in most previous work which largely differ in the methods employed for fitting a model to observations. For example,

recent work by Chang and Zwicker [4] proposed an energy minimization approach to register range scans from articulated objects. For 3D reconstructions based on general non-rigid and articulated image feature trajectories, Paladini *et al.* [15] proposed a non-convex optimization algorithm with projections onto the Stiefel manifold in order to ensure valid camera matrices. The work by Fayad *et al.* [8] phrased the problem of articulated SfM from point correspondences as a discrete-continuous optimization problem which was optimized with an alternating approach switching between discrete graph-cut and continuous optimization. Framing the recovery of an articulated motion also as a non-linear and non-convex optimization problem, Ross *et al.* [16] introduced a probabilistic graphical model for articulated motions. Sturm *et al.*’s model selection based formulation [17] is one of the few approaches, which also considers 1D translational joints. However, more complex restricted motions such as combinations of rotational and translational joints are not addressed. Note that in contrast to the algebraic focus of our work, these approaches put more emphasis on the optimization of the objective function compared to the algebraic constraints due to restricted motions.

Contrary to our analysis of relative transformations, several recent approaches in computer vision are based on subspace intersection constraints [19, 21, 7, 5]. In those approaches, each articulated part has an associated 4D subspace which is given by the span of the trajectories of tracked feature points [18] on that articulated part. As shown in [19, 21], the underlying reason for intersecting trajectory subspaces is again the fixed point assumption due to a joint at a fixed location relative to the two articulated parts. Note that the very same subspace intersection constraints can also be used for motion segmentation purposes [20]. While this already leads to powerful constraints for simple articulated motions around a single fixed joint, the following sections will present an algebraically motivated formulation for the analysis of relative transformations between parts. This formulation clearly goes beyond the current state-of-the-art and leads to novel constraints describing relations between restricted motions. The vectorized representation of relative transformations we will introduce is also related to recent work about rigid factorization-based SfM [2, 3]. This line of work established the relation between trajectory subspaces observed in images and the subspaces spanned by rigid transformations.

3. Notation

The sequence of integers from 1 to F is denoted as $[F] = \{1, \dots, F\}$. Stacking of row or column vectors of the same size into a matrix is denoted with an arrow $\Downarrow_f$, *i.e.* $[\Downarrow_f \mathbf{t}_f^T]$ is the matrix which results by stacking all the vectors $\mathbf{t}_f$ for $f \in [F]$ as rows into a matrix. Matlab indexing notation is used for slicing out subblocks of a matrix, *e.g.* $\mathbf{M}_{:,1:9}$ denotes the first 9 columns of matrix $\mathbf{M}$. A basis for

the orthogonal complement of the columns of a matrix $\mathbf{A}$ will be denoted with $[\mathbf{A}]_{\perp}$, *i.e.* $\mathbf{A}^T [\mathbf{A}]_{\perp} = \mathbf{0}$. The vectorization of a matrix $\mathbf{A}$ is denoted as $\text{vec}(\mathbf{A})$. This operations stacks all the columns of $\mathbf{A}$ below each other in one single column vector, in Matlab notation $\text{vec}(\mathbf{A}) = \mathbf{A}(:)$. The representation of rotations in angle-axis notation will be very important throughout this paper. Specifically, the rotation matrix around an axis $\mathbf{a}$ by an angle α can be represented in the following way

$$\mathbf{R}_{\mathbf{a},\alpha} = \cos \alpha \mathbf{I}_3 + (1 - \cos \alpha) \mathbf{a} \mathbf{a}^T + \sin \alpha [\mathbf{a}]_{\times}, \quad (1)$$

where $[\mathbf{a}]_{\times}$ denotes the cross-product matrix, *i.e.* $[\mathbf{a}]_{\times} \mathbf{b}$ denotes the cross-product between the vectors $\mathbf{a}$ and $\mathbf{b}$.

4. Rigid Motion Subspaces

As an element of SE_3 , the special Euclidean group in $\mathbb{R}^3$, a rigid transformation at frame $f \in [F]$

$$\mathcal{T}_f = \begin{bmatrix} \mathbf{R}_f & \mathbf{t}_f \\ \mathbf{0}^T & 1 \end{bmatrix} \quad (2)$$

has 3 non-linear and 3 linear degrees of freedom. This highly non-linear special Euclidean group can be embedded in a higher-dimensional *linear* space and we are going to see that this renders the analysis of restricted motions particularly simple. Hence, rigid transformations will be considered as points in $\mathbb{R}^{13}$, *i.e.* $(\text{vec}(\mathbf{R}_f)^T, \mathbf{t}_f^T, 1)$ and F general rigid motions span a 12D affine subspace embedded in $\mathbb{R}^F$ which is spanned by the columns of the matrix

$$\left[\downarrow_f \left(\text{vec}(\mathbf{R}_f)^T, \mathbf{t}_f^T, 1 \right) \right]. \quad (3)$$

This representation of rigid motions as an affine subspace in $\mathbb{R}^F$ has already appeared in [2]. However, that work focused on factorization-based structure-from-motion and the major challenge was to find the matrix in Eq. (3) with a factorization algorithm, *i.e.* the recovery of these rigid transformations based on factorization of feature trajectory data. In contrast, the present work analyses restricted motions and assumes that the relative transformations $\mathcal{T}_f$ at each frame $f \in [F]$ between two parts are given as input. The goal is then to extract all the aspects of the restricted motion by analyzing the subspace structure of the matrix in Eq. (3). These aspects include the determination of the type of articulated or restricted motion and all its parameters, *e.g.* the orientation and location of rotation axes amongst others. Such an analysis can actually be slightly simplified by considering $\mathcal{T}_f - \mathbf{I}_4$ instead of $\mathcal{T}_f$ since the 'homogeneous 1' cancels in that case and this allows to reason about 12D *linear* subspaces embedded in $\mathbb{R}^F$. Hence, rigid transformations will be mapped to vectors $(\text{vec}(\mathbf{R}_f - \mathbf{I}_3)^T, \mathbf{t}_f^T) \in \mathbb{R}^{12}$ and F general rigid motions will therefore span a 12-dimensional linear subspace embedded in $\mathbb{R}^F$ which is defined by the column span of the *motion matrix*

$$\mathbf{M} = \left[\downarrow_f \left(\text{vec}(\mathbf{R}_f - \mathbf{I}_3)^T, \mathbf{t}_f^T \right) \right] \in \mathbb{R}^{F \times 12}. \quad (4)$$

5. Motion Signatures

The important observation is that restricted motions generally do not entirely span the aforementioned 12D space. Indeed, in the following we are going to show that each type of restricted motion will yield a specific low-dimensional subspace structure which allows to distinguish between different motions by just considering the matrix $\mathbf{M}$. More specifically, we propose a tuple of integers called *signature* in order to capture the low-dimensional subspace structure of an restricted motion. This signature is defined as a function of the motion matrix $\mathbf{M}$ in the following way

$$\text{sig}(\mathbf{M}) = (\text{rank}(\mathbf{M}_{:,1:9}), \text{rank}(\mathbf{M}) - \text{rank}(\mathbf{M}_{:,1:9})).$$

The signature $\text{sig}(\mathbf{M})$ uniquely determines the type of restricted motion encoded in $\mathbf{M}$ (see list in Tab. 1). The first entry of the signature $\text{sig}(\mathbf{M}) = (r, d)$ entirely determines the number of fixed rotation axes involved in the restricted motion whereas the second entry specifies the dimensionality d of the subspace in which the object translates over time. Sec. 7 will explain in more detail what we mean with translations in restricted motions. Tab. 1 summarizes the results of the analysis based on our unified framework for articulated and restricted motions whereas Tab. 2 shows several intuitive examples of specific instances of such motions handled by our framework. Moreover, once the signature is computed and the type of restricted motion is thereby determined, all aspects of this motion can be directly extracted by carefully analyzing the nullspace structure of $\mathbf{M}$ (see Sec. 8).

It will turn out to be important to distinguish between two subspaces determined by the matrix $\mathbf{M}$. The translation subspace is spanned by the columns of the translational part $\mathbf{M}_{:,10:12}$ whereas the rotation subspace is spanned by the rotational part of the motion matrix $\mathbf{M}_{:,1:9}$. Note that the second entry of the signature encodes the dimensionality d of the subspace of the translation subspace which is not yet contained in the rotation subspace. It is not obvious why the translation subspace is entirely contained in the rotation subspace in the absence of any dynamic translation, *i.e.* when $d = 0$. This is especially true for rotations around two fixed axes. The following two sections will explain why this is the case. Initially in Sec. 6, time-varying translations will be ignored, *i.e.* we assume $d = 0$ and once an appropriate formulation has been derived, we will treat the case $d \neq 0$ in Sec. 7. As the case of pure translations is trivial (*i.e.* $r = 0$), we omit any further discussion of this case.

6. Non-Translating Joints

In the following, the signature for rotations around one, two, or three fixed axes will be explained under the assumption that the joint is not translating, *i.e.* $d = 0$. Specifically, we will derive the values of the two entries of the signature for each type of restricted motion. Two questions need to be addressed in order to do so. Firstly, the dimensionality of the rotation subspace $\text{span}(\mathbf{M}_{:,1:9})$ needs to be derived.

Type of Articulation / Restricted Motion	Signature	Formula for relative transformation $\mathcal{T}_f$ with $\mathbf{T} \in \mathbb{R}^{3 \times d}$
Pure d -dimensional translation	$(0, d)$	$\begin{bmatrix} \mathbf{I}_3 & \mathbf{T}\tilde{\mathbf{t}}_f \\ \mathbf{0}_{1 \times 3} & 1 \end{bmatrix}$
Rotation around one axis and d -dimensional translation	$(2, d)$	$\begin{bmatrix} \mathbf{R}_{\mathbf{a}, \alpha_f} & \mathbf{t} - \mathbf{R}_{\mathbf{a}, \alpha_f} \mathbf{t} + \mathbf{T}\tilde{\mathbf{t}}_f \\ \mathbf{0}_{1 \times 3} & 1 \end{bmatrix}$
Rotation around two axes and d -dimensional translation	$(8, d)$	$\begin{bmatrix} \mathbf{R}_{\mathbf{b}, \beta_f} & \mathbf{t}_b + \mathbf{T}\tilde{\mathbf{t}}_f \\ \mathbf{0}_{1 \times 3} & 1 \end{bmatrix} \begin{bmatrix} \mathbf{I}_3 & -\mathbf{t}_b + \mathbf{t}_a \\ \mathbf{0}_{1 \times 3} & 1 \end{bmatrix} \begin{bmatrix} \mathbf{R}_{\mathbf{a}, \alpha_f} & -\mathbf{t}_a \\ \mathbf{0}_{1 \times 3} & 1 \end{bmatrix}$
Arbitrary rotations and d -dimensional translation	$(9, d)$	$\begin{bmatrix} \mathbf{R}_f & \mathbf{t} - \mathbf{R}_f \mathbf{t} + \mathbf{T}\tilde{\mathbf{t}}_f \\ \mathbf{0}_{1 \times 3} & 1 \end{bmatrix}$

Table 1: List of possible signatures: The dimension d of the dynamic translations $\mathbf{T}\tilde{\mathbf{t}}_f$ can be zero. Rotations around two intersecting axes are captured by choosing $\mathbf{t}_b = \mathbf{t}_a$. Note that the axes $\mathbf{a}$ and $\mathbf{b}$ do not have to be orthogonal. The important class of planar motions corresponds to signature $(2, 2)$ with a rotation axis orthogonal to the translation direction, *i.e.* $\mathbf{a}^T \mathbf{T} = \mathbf{0}^T$.

Secondly, we need to show that the translation subspace span $(\mathbf{M}_{:,10:12})$ is entirely contained inside the rotation subspace since a signature of the form $(r, 0)$ encodes exactly this property. In preparation for translating joints in Sec. 7, this property will be shown by finding a matrix $\mathbf{X} \in \mathbb{R}^{9 \times 3}$ such that

$$\mathbf{M} \begin{bmatrix} \mathbf{X} \\ \mathbf{I}_3 \end{bmatrix} = \mathbf{0}_{F \times 3}. \quad (5)$$

6.1. Rotations Around One Axis

A rotation around an axis $\mathbf{a}$ which is located at point $\mathbf{t}$ by a time-varying angle α_f reads like

$$\begin{bmatrix} \mathbf{R}_f & \mathbf{t}_f \\ \mathbf{0}^T & 1 \end{bmatrix} = \begin{bmatrix} \mathbf{R}_{\mathbf{a}, \alpha_f} & \mathbf{t} - \mathbf{R}_{\mathbf{a}, \alpha_f} \mathbf{t} \\ \mathbf{0}^T & 1 \end{bmatrix}, \quad (6)$$

which can be understood as changing the coordinate system by first moving the location of the axis to the origin, then applying the rotation, before finally reestablishing the original coordinate origin. Let us have a look at the axis-angle representation of

$$\mathbf{R}_{\mathbf{a}, \alpha_f} - \mathbf{I}_3 = (1 - \cos \alpha_f) (\mathbf{a} \mathbf{a}^T - \mathbf{I}_3) + \sin \alpha_f [\mathbf{a}]_{\times}. \quad (7)$$

This shows that the rotation subspace will only span a 2-dimensional subspace determined by the column span of $[\downarrow_f (1 - \cos \alpha_f, \sin \alpha_f)] \in \mathbb{R}^{F \times 2}$ since

$$\begin{aligned} \mathbf{M}_{:,1:9} &= [\downarrow_f \text{vec}(\mathbf{R}_{\mathbf{a}, \alpha_f} - \mathbf{I}_3)]^T \\ &= [\downarrow_f (1 - \cos \alpha_f, \sin \alpha_f)] \begin{bmatrix} \text{vec}(\mathbf{a} \mathbf{a}^T - \mathbf{I}_3)^T \\ \text{vec}([\mathbf{a}]_{\times})^T \end{bmatrix}. \end{aligned} \quad (8)$$

This explains the first entry of the signature in the case of rotations around one axis.

What remains to be shown is that the translation subspace is entirely contained in the rotation subspace. It is well-known, that the location of a non-translating joint is unaffected by a rotation around this joint and hence $\mathbf{t}$ represents a fixed point. In turn, we have

$$([\mathbf{R}_{\mathbf{a}, \alpha_f}, \mathbf{t} - \mathbf{R}_{\mathbf{a}, \alpha_f} \mathbf{t}] - [\mathbf{I}_3 \quad \mathbf{0}]) \begin{pmatrix} \mathbf{t} \\ 1 \end{pmatrix} = \mathbf{0}, \quad (9)$$

and application of a property of the Kronecker-product¹

¹For any matrices of compatible size, the following holds:

yields the equation $\mathbf{M}[\mathbf{X}^T, \mathbf{I}_3]^T = \mathbf{0}_{F \times 3}$ with $\mathbf{X} = \mathbf{t} \otimes \mathbf{I}_3$. This shows in particular that the translation subspace span $(\mathbf{M}_{:,10:12})$ is a linear combination of the rotation subspace span $(\mathbf{M}_{:,1:9})$ and hence entirely contained in the latter.

6.2. Rotations Around Two Axes

The sequential application of rotations around two fixed axes $\mathbf{a}$ at location $\mathbf{t}_a$ and $\mathbf{b}$ at location $\mathbf{t}_b$ which are not necessarily located at the same point in space looks like

$$\begin{bmatrix} \mathbf{R}_f & \mathbf{t}_f \\ \mathbf{0}^T & 1 \end{bmatrix} = \begin{bmatrix} \mathbf{R}_{\mathbf{b}, \beta_f} & \mathbf{t}_b - \mathbf{R}_{\mathbf{b}, \beta_f} \mathbf{t}_b \\ \mathbf{0}^T & 1 \end{bmatrix} \begin{bmatrix} \mathbf{R}_{\mathbf{a}, \alpha_f} & \mathbf{t}_a - \mathbf{R}_{\mathbf{a}, \alpha_f} \mathbf{t}_a \\ \mathbf{0}^T & 1 \end{bmatrix}.$$

Since $\mathbf{a}$ and $\mathbf{b}$ are eigenvectors of $\mathbf{R}_{\mathbf{a}, \alpha_f}$ and $\mathbf{R}_{\mathbf{b}, \beta_f}$ respectively, it holds that $\mathbf{b}^T \mathbf{R}_{\mathbf{b}, \beta_f} \mathbf{R}_{\mathbf{a}, \alpha_f} \mathbf{a} = \mathbf{b}^T \mathbf{a}$ is constant over time. Hence, we have that $\mathbf{M}_{:,1:9} (\mathbf{a} \otimes \mathbf{b}) = \mathbf{0}_{F \times 1}$ which shows that in this case the rank is at most 8. Using the angle-axis representation, it can actually be shown that the rank equals 8.

The translations are equal to $\mathbf{t}_f = \mathbf{t}_b + \mathbf{R}_{\mathbf{b}, \beta_f} (-\mathbf{t}_b + \mathbf{t}_a - \mathbf{R}_{\mathbf{a}, \alpha_f} \mathbf{t}_a)$ and therefore

$$\begin{aligned} \mathbf{M}_{:,10:12} &= [\downarrow_f - \text{vec}(\mathbf{R}_{\mathbf{b}, \beta_f} - \mathbf{I}_3)]^T [\mathbf{t}_b \otimes \mathbf{I}_3] \\ &\quad - \text{vec}(\mathbf{R}_{\mathbf{b}, \beta_f} (\mathbf{R}_{\mathbf{a}, \alpha_f} - \mathbf{I}_3))^T [\mathbf{t}_a \otimes \mathbf{I}_3]. \end{aligned} \quad (10)$$

It can be shown (*e.g.* by inserting the angle-axis representation for rotation matrices) that the column span of both $[\downarrow_f - \text{vec}(\mathbf{R}_{\mathbf{b}, \beta_f} - \mathbf{I}_3)]^T$ and $[\downarrow_f - \text{vec}(\mathbf{R}_{\mathbf{b}, \beta_f} (\mathbf{R}_{\mathbf{a}, \alpha_f} - \mathbf{I}_3))^T]$ are contained in the column span of $\mathbf{M}_{:,1:9}$. Therefore, there again exists a matrix $\mathbf{X}$ such that Eq. (5) holds. We refer to the supplemental material [11] for a detailed derivation and an explicit formula for $\mathbf{X}$ since no valuable insight can be gained by the required algebraic manipulations.

6.3. Rotations Around Three Axes

Here, we assume a ball joint, *i.e.* all three axes are located at the same point in space $\mathbf{t}$. Rotations around three fixed axes are obviously fully general, *i.e.* any rotation can be

$$\text{vec}(\mathbf{A} \mathbf{Y} \mathbf{B}) = [\mathbf{B}^T \otimes \mathbf{A}] \text{vec}(\mathbf{Y}).$$

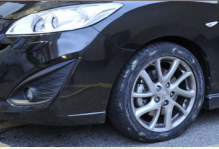
Example	Description	Signature	Additional Constraints
	Planar motion, <i>i.e.</i> a rigid object which translates in a plane orthogonal to its rotation axis.	(2, 2)	Rotation axis and translation direction are orthogonal $\mathbf{a}^T \mathbf{T} = \mathbf{0}^T$
	A wheel rolling on the ground: the wheel rotates around an axis with fixed orientation and translates at the same time along a line orthogonal to the axis.	(2, 1)	Rotation axis and translation direction are orthogonal, <i>i.e.</i> $\mathbf{a}^T \mathbf{T} = 0$, and linear relationship between α_f and magnitude of translation $\mathbf{t}_f$.
	A rotatable blackboard: the stand undergoes a planar motion whereas the blackboard itself also rotates around an axis which is parallel to the plane of planar motion.	(8, 2)	Intersecting axes, <i>i.e.</i> $\mathbf{t}_a = \mathbf{t}_b$, and orthogonal axes $\mathbf{a}^T \mathbf{b} = 0$.
	Drawer opening and closing along a 1D line.	(0, 1)	

Table 2: Examples of restricted motions which can be handled by the formulation presented in this paper. Note that a static camera observing the dynamic object can act itself as one of the two moving parts defining the relative motion between the two parts.

decomposed into three sequential rotations around these three fixed axes. Hence the rotation subspace spans a 9D space. Moreover, since $\mathbf{t}$ is again unaffected by the rotation, the same derivation as in Sec. 6.1 applies and hence Eq. (5) holds with $\mathbf{X} = \mathbf{t} \otimes \mathbf{I}_3$.

7. Translating Joints

Modeling also a time-varying translational component $\hat{\mathbf{t}}_f$ in the restricted motion between two parts leads to slightly more complex formulas. For example, a rotation around one axis which at the same time translates is modeled as

$$\begin{bmatrix} \mathbf{R}_f & \mathbf{t}_f \\ \mathbf{0}^T & 1 \end{bmatrix} = \begin{bmatrix} \mathbf{I}_3 & \hat{\mathbf{t}}_f \\ \mathbf{0}^T & 1 \end{bmatrix} \begin{bmatrix} \mathbf{R}_{\mathbf{a}, \alpha_f} & \mathbf{t} - \mathbf{R}_{\mathbf{a}, \alpha_f} \mathbf{t} \\ \mathbf{0}^T & 1 \end{bmatrix}. \quad (11)$$

Let us assume that the time-varying translations are restricted to a d -dimensional subspace, *i.e.* $\hat{\mathbf{t}}_f = \mathbf{T} \tilde{\mathbf{t}}_f$ with $\mathbf{T} \in \mathbb{R}^{3 \times d}$. Given the observed relative transformations $[\mathbf{R}_f, \mathbf{t}_f]$, we would like to extract not only the parameters of the rotation (*e.g.* the axes of rotation and their locations) but also the low-dimensional subspace spanned by the translations, *i.e.* $\text{span}(\mathbf{T})$.

Coming back to the previous example, note that neither $\mathbf{t} + \hat{\mathbf{t}}_f - \mathbf{R}_{\mathbf{a}, \alpha_f} \mathbf{t}$ nor the translation of the inverse relative transformation $-\mathbf{R}_{\mathbf{a}, \alpha_f}^T (\mathbf{t} + \hat{\mathbf{t}}_f) + \mathbf{t}$ has a low-dimensional structure, even though $\hat{\mathbf{t}}_f$ might have. If we knew the location $\mathbf{t}$ of the joint, we could apply a coordinate transformation to set the origin of the coordinate system to this joint location. In that case, $\mathbf{t}$ becomes equal to the zero-vector and the low-dimensional subspace could directly be read off from the relative transformation. However, the joint location is usually not known upfront. Nevertheless, the translational component of the relative transformation can be reformulated according to $[\mathbf{R}_f - \mathbf{I}_3] \mathbf{t} + \mathbf{t}_f = \hat{\mathbf{t}}_f$. If the $\hat{\mathbf{t}}_f$ originate from a d -dimensional subspace spanned by the columns of

$\mathbf{T} \in \mathbb{R}^{3 \times d}$, we get

$$[\Rightarrow_f [\mathbf{R}_f - \mathbf{I}_3] \mathbf{t} + \mathbf{t}_f] = [\Rightarrow_f \hat{\mathbf{t}}_f] = \mathbf{T} [\Rightarrow_f \tilde{\mathbf{t}}_f]. \quad (12)$$

One approach to solve this low-rank problem is to use polynomial minor constraints in the entries of $\mathbf{t}$. This will lead to trivariate polynomial equations of grade $d + 1$. Solving such multivariate polynomial systems is a delicate issue. Therefore, here we sidestep this by building upon the previously derived formula in Eq. (5). This has the additional benefit that a single unified framework can handle all the cases mentioned in Tab. 1. As can be seen by the example in Eq. (11), the time-varying translations $\hat{\mathbf{t}}_f$ are simply added as additional offsets to the translations which would result without time-varying translations. Hence, using the matrices $\mathbf{X} \in \mathbb{R}^{9 \times 3}$ derived in Sec. 6, we see that

$$\mathbf{M} \begin{bmatrix} \mathbf{X} \\ \mathbf{I}_3 \end{bmatrix} = [\Downarrow_f \hat{\mathbf{t}}_f^T] = [\Downarrow_f \tilde{\mathbf{t}}_f^T] \mathbf{T}^T. \quad (13)$$

The component of the translations $\mathbf{M}_{:,10:12}$ which is linearly independent of the rotational part $\mathbf{M}_{:,1:9}$ is therefore entirely due to the dynamic translation. Hence, the second entry of the signature equals d because the dynamic translations are restricted to a d -dimensional subspace. Finally, since the matrix $\mathbf{T}$ is constant throughout time we finally arrive at the following equation which will turn out to be crucial for the extraction of the motion parameters (see Sec. 8.2)

$$\mathbf{M} \begin{bmatrix} \mathbf{X} \\ \mathbf{I}_3 \end{bmatrix} [\mathbf{T}]_{\perp} = [\Downarrow_f \tilde{\mathbf{t}}_f^T] \mathbf{T}^T [\mathbf{T}]_{\perp} = \mathbf{0}_{F \times 3-d}. \quad (14)$$

Remark: Planar translations or translations along a line are a relative concept. Specifically, considering the inverse relative transformations $\mathcal{T}_f^{-1}$ might not reveal a low-dimensional structure $\mathbf{T} \in \mathbb{R}^{3 \times d}$ contained in the translations, even though $\mathcal{T}_f$ contains one. Therefore, in order to be on the safe side, both sets of transformations need to be analyzed.

8. Extraction of Parameters

Having shown that the signature provides an unique pattern for varying types of restricted motion, an appropriate method to extract the motion parameters can be chosen according to the signature encoded in the motion matrix $\mathbf{M}$. In the following, we shortly outline how these parameters can be extracted by simply solving several linear systems.

8.1. Rotation Axes and Angles

If the signature tells us that the motion is around one fixed axis (*i.e.* the first entry of the signature equals 2), the linear system $\mathbf{M}_{:,1:9}(\mathbf{a} \otimes \mathbf{I}_3) = \mathbf{0}_{F \times 3}$ needs to be solved for the axis $\mathbf{a}$. This is based on the fact that the axis $\mathbf{a}$ is an eigenvector with eigenvalue 1 of each $\mathbf{R}_{\mathbf{a},\alpha_f}$. The approach for rotations around two fixed axes is based on the fact that $\mathbf{b}^T \mathbf{R}_{\mathbf{b},\beta_f} \mathbf{R}_{\mathbf{a},\alpha_f} \mathbf{a} = \mathbf{b}^T \mathbf{a}$ which is constant throughout time. Hence, the only nullvector of $\mathbf{M}_{:,1:9}$ must be of the form $\mathbf{n} = \mathbf{a} \otimes \mathbf{b}$. Hence, in case the signature predicts a rotation around two axes (*i.e.* the first entry equals 8), the one-dimensional nullspace $\mathbf{n} \in \mathbb{R}^9$ of $\mathbf{M}_{:,1:9}$ is computed first, this nullspace is then reshaped into a 3-by-3 matrix $\mathbf{N}$, and lastly a rank-1 decomposition of this reshaped matrix reveals the two axes $\mathbf{N} = \mathbf{b}\mathbf{a}^T$. The extraction of the angles is based on Eq. (1) for $\mathbf{R}_{\mathbf{a},\alpha}$ which leads to the following formulas for any vector $\mathbf{c}$

$$\mathbf{a}^T [\mathbf{c}]_{\times} \mathbf{R}_{\mathbf{a},\alpha} \mathbf{c} = \sin \alpha (1 - (\mathbf{c}^T \mathbf{a})^2) \quad (15)$$

$$\mathbf{c}^T \mathbf{R}_{\mathbf{a},\alpha} \mathbf{c} = \cos \alpha (1 - (\mathbf{c}^T \mathbf{a})^2) + (\mathbf{c}^T \mathbf{a})^2. \quad (16)$$

For the extraction of α_f in the single axis motion case, Eq. (15) and Eq. (16) directly yield $\cos \alpha_f$ and $\sin \alpha_f$ respectively if the vector $\mathbf{c}$ is chosen orthogonal to the axis $\mathbf{a}$. For the two-axes case, the vector $\mathbf{c}$ is chosen as $\mathbf{a}$ and $\mathbf{b}$ giving, since $\mathbf{R}_f = \mathbf{R}_{\mathbf{b},\beta_f} \mathbf{R}_{\mathbf{a},\alpha_f}$:

$$\mathbf{b}^T \mathbf{R}_f \mathbf{b} = \cos \alpha_f (1 - (\mathbf{b}^T \mathbf{a})^2) + (\mathbf{b}^T \mathbf{a})^2 \quad (17)$$

$$\mathbf{a}^T \mathbf{R}_f \mathbf{a} = \cos \beta_f (1 - (\mathbf{a}^T \mathbf{b})^2) + (\mathbf{a}^T \mathbf{b})^2 \quad (18)$$

$$\mathbf{b}^T [\mathbf{a}]_{\times} \mathbf{R}_f \mathbf{a} = \sin \beta_f (1 - (\mathbf{a}^T \mathbf{b})^2) \quad (19)$$

$$\mathbf{a}^T [\mathbf{b}]_{\times} \mathbf{R}_f^T \mathbf{b} = -\sin \alpha_f (1 - (\mathbf{b}^T \mathbf{a})^2), \quad (20)$$

from which the angles α_f and β_f can be readily extracted.

8.2. Translations

The extraction of the translational parameters is based on our derivations in Sec. 7. Especially Eq. (14) is important in so far as it allows to recover a basis $\mathbf{T}$ for the time-varying translations by analyzing the nullspace of $\mathbf{M}$. More specifically, we compute the nullspace $\mathbf{N}$ of $\mathbf{M}$, *i.e.* $\mathbf{M}\mathbf{N} = \mathbf{0}$, *e.g.* by a singular value decomposition of $\mathbf{M}$. Then the column span of the columns of $\mathbf{N}$ restricted to the last three entries encode a basis for the orthogonal complement $[\mathbf{T}]_{\perp}$, *i.e.* $\text{span}(\mathbf{N}_{10:12,:}) = \text{span}([\mathbf{T}]_{\perp})$. This orthogonal complement completely determines the subspace of dynamic translations. Note that the dimensionality of the nullspace of $\mathbf{M}$ might be higher than the dimensionality $3 - d$ of the orthogonal complement $[\mathbf{T}]_{\perp}$ since the rotational part $\mathbf{M}_{:,1:9}$ might

be rank-deficient, as well. However, the part of the nullspace due to the rotational part does not interfere with the last three rows of $\mathbf{N}$ which are entirely determined by $[\mathbf{T}]_{\perp}$. Nevertheless, the nullspace due to the rotational part alone could be eliminated by first computing a rank-revealing singular value decomposition of $\mathbf{M}_{:,1:9} = \mathbf{U}\Sigma\mathbf{V}^T$ and then computing the nullspace $\mathbf{N} \in \mathbb{R}^{12 \times 3-d}$ of $[\mathbf{U}, \mathbf{M}_{:,10:12}]$ whose dimensionality will then equal the dimensionality of $\text{span}([\mathbf{T}]_{\perp})$. Once $\mathbf{T}$ is known, the computation of a decomposition of the observed translations $\mathbf{t}_f$ into more semantically meaningful parts boils down to a linear system of equations. For example, the translations observed by rotations around two axes equal $\mathbf{t}_f = \mathbf{t}_b + \mathbf{R}_{\mathbf{b},\beta_f}(-\mathbf{t}_b + \mathbf{t}_a - \mathbf{R}_{\mathbf{a},\alpha_f} \mathbf{t}_a) + \mathbf{T}\tilde{\mathbf{t}}_f$, which are linear in the unknowns $\mathbf{t}_a$, $\mathbf{t}_b$, and $\tilde{\mathbf{t}}_f$. Given sufficiently many frames, these unknowns can be computed in the least-squares sense, for example. Actually, since $\mathbf{t}_a$ and $\mathbf{t}_b$ are only defined up to translations along $\mathbf{a}$ and $\mathbf{b}$, respectively, the locations of the axes can be parametrized by $\mathbf{t}_a = [\mathbf{a}]_{\perp} \tilde{\mathbf{t}}_a$ and $\mathbf{t}_b = [\mathbf{b}]_{\perp} \tilde{\mathbf{t}}_b$.

Note that there are some instances where the location of the axis not well-defined. Specifically, the location $\mathbf{t}_a$ of an axis $\mathbf{a}$ is not uniquely defined if $\mathbf{a}^T \mathbf{T} = \mathbf{0}$ and $d = 2$ since in that case, shifting the axis inside the plane $\mathbf{t}_a + \mathbf{T}\mathbf{t}_0$ can be compensated by the time varying $\mathbf{T}\tilde{\mathbf{t}}_f$. The important class of planar motions belong to this category. Nevertheless, such cases do not pose a problem for our algorithm, the system matrix of the resulting linear system will simply have a 2-dimensional nullspace.

9. Experiments

We performed multiple experiments with image data from challenging motion sequences. Since our method is entirely based on relative transformations between two parts, let us quickly explain how we extracted these transformations from pure image data. Note that any other method to extract such transformations could be used, as well (*e.g.* a motion capture system or ICP on point clouds from measurements obtained by a depth camera). We build on top of standard rigid SfM [10] with intrinsically calibrated cameras. Feature points are extracted and matched across different frames and a robust RANSAC [9] stage extracts rigid transformations from two- and three-view relations. In case of a static camera observing an object which undergoes a restricted motion (experiments in Sec. 9.2 and Sec. 9.3), the camera actually acts as the second part. Hence, the SfM pipeline will report a single transformation between pairs of frames which corresponds to the relative motion of the two parts between these two frames. In case of a moving camera (experiment in Sec. 9.1), an initial 3D model is reconstructed from two or more views assuming that the configuration of the articulated or otherwise motion-restricted object stays fixed during those views thereby also fixing the scale ambiguity between parts. Sequential 3-point RANSAC is then used to extract all rigid



Figure 2: Our method accurately recovers the rotation axis $\mathbf{a}$ (in red) of a laptop opening and closing its screen from pure imagery data taken by a moving camera. This articulated motion has signature $(2, 0)$ and is thus equivalent to the motion of a door, for example.

transformations with sufficient support between this initial model and new views. The so obtained transformations are split into disjoint groups based on the spatial distance between their inlier sets. Each group then corresponds to the motion of one part. Lastly, the relative motion between two parts can be recovered from the two corresponding groups by expressing their motions w.r.t. to one fixed frame. This basically factors out the motion of the camera and only the motions of the parts remain.

9.1. Hinge Joint

A sequence of 23 images of a laptop opening and closing the screen were taken with a moving camera. Features were tracked on both, the laptop base and the screen. Since this motion corresponds to a rotation around one single axis with fixed location, the signature equals $(2, 0)$. Being an articulation type which has already been studied extensively in previous work, here we would like to show that our general framework flawlessly handles this type, as well. Despite a moving camera, we are able to accurately compute the orientation $\mathbf{a}$ and location $\mathbf{t}$ of the rotation axis based on our method (see Fig. 2). The extracted articulated motion parameters can be used to generate novel, unobserved configurations, as demonstrated in the supplemental material [11].

9.2. Planar Motion and Two-Axes Rotations

For this experiment, 27 image of a moving blackboard were taken with a static camera. Features were matched exclusively on the black writing area. These features undergo a complex motion since the blackboard stands on wheels and can therefore be rotated and translated according to a planar motion. On top of that, the black writing area can be rotated around a horizontal axis leading to a restricted motion with signature $(8, 2)$ which previous approaches could not handle. Our method can successfully extract the two rotation axes $\mathbf{a}$ and $\mathbf{b}$ and the two-dimensional span of the translation directions $\mathbf{T} \in \mathbb{R}^{3 \times 2}$. Due to one part of the motion being a planar motion, the location $\mathbf{t}_b$ of the axis $\mathbf{b}$ is not defined. Note that this is not a defect of our method but rather an inherent property of planar motions. As Fig. 3 shows, the well-defined location $\mathbf{t}_a$ of the first axis $\mathbf{a}$ is recovered correctly. Furthermore, even though our method does not explicitly enforce orthogonalities contained in the data such as between the rotation axes $\mathbf{a}^T \mathbf{b} = 0$ or between

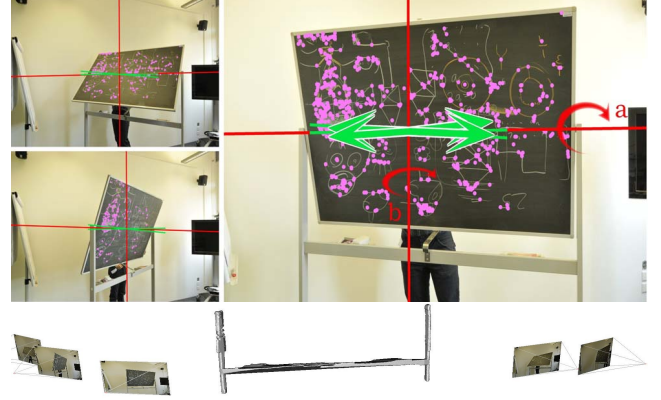


Figure 3: *Top*: The restricted motion parameters of a moving blackboard undergoing a rotation around a horizontal axis $\mathbf{a}$ (in red) followed by a planar motion are recovered accurately by our approach. The two extracted rotation axes are shown in red and the translation directions $\mathbf{T}$ (parallel to the floor) in green. Note that the location $\mathbf{t}_a$ of the axis $\mathbf{a}$ is well-defined (and recovered), whereas the location $\mathbf{t}_b$ of axis $\mathbf{b}$ is not defined due to the planar motion. For visualisation purposes, we chose $\mathbf{t}_b$ at the center of mass of the tracked points (visualized in pink).

Bottom: As discussed in Sec. 9.2, our approach enables a visual-hull reconstruction of the blackboard stand, even though no feature points on the stand have been tracked (i.e. no poses were known relative to it).

the translations and the second axis $\mathbf{b}^T \mathbf{T} = \mathbf{0}^T$, we recover angles $\angle(\mathbf{a}, \mathbf{b}) = 90.0345^\circ$ and $\angle(\mathbf{b}, [\mathbf{T}]_\perp) = 0.0510^\circ$, a clear indicator for the accuracy of our estimates. As a further application, once the two rotations are disentangled, we can compute the intermediate motion of a putative part as would be observed without the rotation $\mathbf{R}_{\mathbf{a}, \alpha_f}$, i.e. in the present case without the rotation around the horizontal axis. This recovered intermediate motion together with silhouette images obtained with background subtraction permits for example the computation of a visual hull of the blackboard stand, as shown in Fig. 3. Remember that no points could be tracked on the stand and its shape is only deduced based on its silhouette consistency w.r.t this recovered intermediate motion.

9.3. Rotation Around a Translating Axis

This experiment is based on the motion of the front wheel of a car rolling on a straight line on the street. Features on the wheel are extracted and matched across 21 images taken by a static camera. While this motion has similarity to a hinge joint, the one-dimensional dynamic translation leads to a signature $(2, 1)$ which makes this particular restricted motion a hard instance. We refer to Fig. 1 and Fig. 4 for a visualization of the extracted motion parameters.

Having recovered the motion parameters including the time-varying rotation angles α_f and translations $\mathbf{T} \tilde{\mathbf{t}}_f$, we can check for a linear relation $\tilde{\mathbf{t}}_f = \alpha_f R$ between α_f and $\tilde{\mathbf{t}}_f$ to recover the radius R of the wheel and its contact line with the street. As can be seen in Fig. 4, a linear model clearly relates these estimates in this data set.

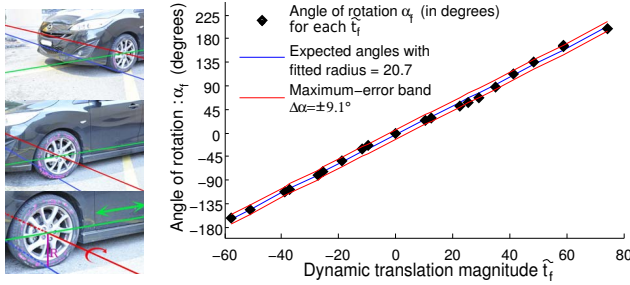


Figure 4: A wheel of a car rolling on a straight line on the street corresponds to a signature $(2, 1)$. The tracked feature points are shown in pink, the extracted rotation axis in red, the translation direction in green, and the contact line in blue. Once the motion parameters are recovered, the angles α_f and translation magnitudes $\tilde{t}_f$ can be explored, looking for potential dependencies. Here a linear relation is found which actually shows the radius of the wheel (in arbitrary unit, since SfM reconstruction is up to scale).

10. Conclusions and Future Work

In this paper, we have analysed the relative transformation between two motion-restricted parts. The vectorized relative transformations gave rise to a motion matrix. A signature can be computed from the restricted motion subspace induced by this motion matrix which exactly specifies the type of restricted motion. Together with a careful analysis of the nullspace-structure of the motion matrix, this leads to a general framework for articulated and restricted motions between two parts. The framework has been successfully applied to several challenging data sets showcasing how existing and novel restricted motion types can be handled in the same way.

In future work, we will address the issue of assigning detected rigid transformations to individual parts. As described in Sec. 9, we currently rely on a static camera or on a simple heuristic based on the spatial distance between the points in the inlier sets of the transformations found by the SfM pipeline. Furthermore, we are investigating robust model selection / rank-detections for the singular values of the motion matrix since SfM can return erroneous relative transformations, especially for nearly degenerate or ill-condition motion sequences.

Acknowledgments

This work was supported by the 4DVideo ERC Starting Grant Nr. 210806 and by the Swiss National Science Foundation under Project Nr. 143422.

R. Angst is a recipient of the Google Europe Fellowship in Computer Vision, and this research is supported in part by this Google Fellowship.

References

- [1] S. Agarwal, N. Snavely, I. Simon, S. M. Seitz, and R. Szeliski. Building rome in a day. In *ICCV*, 2009.
- [2] R. Angst and M. Pollefeys. Static multi-camera factorization using rigid motion. In *ICCV*, 2009.
- [3] R. Angst and M. Pollefeys. 5d motion subspaces for planar motions. In *ECCV*, 2010.
- [4] W. Chang and M. Zwicker. Global registration of dynamic range scans for articulated model reconstruction. *ACM Transactions on Graphics*, 2011.
- [5] N. da Silva and J. Costeira. Subspace segmentation with outliers: A grassmannian approach to the maximum consensus subspace. In *CVPR*, 2008.
- [6] E. de Aguiar, C. Theobalt, and H.-P. Seidel. Automatic learning of articulated skeletons from 3d marker trajectories. In *ISVC*, 2006.
- [7] E. Elhamifar and R. Vidal. Sparse subspace clustering: Algorithm, theory, and applications. *CoRR*, abs/1203.1005, 2012.
- [8] J. Fayad, C. Russell, and L. de Agapito. Automated articulated structure and 3d shape recovery from point correspondences. In *ICCV*, 2011.
- [9] M. A. Fischler and R. C. Bolles. Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography. *Communications of the ACM*, pages 381–395, 1981.
- [10] R. Hartley and A. Zisserman. *Multiple view geometry in computer vision*. Cambridge University Press, 2004.
- [11] B. Jacquet, R. Angst, and M. Pollefeys. Articulated and restricted motion subspaces and their signatures : Supplemental material. In *CVPR*, 2013.
- [12] D. Katz and O. Brock. Interactive segmentation of articulated objects in 3d. In *Workshop on Mobile Manipulation at ICRA*, 2011.
- [13] A. G. Kirk, J. F. O’Brien, and D. A. Forsyth. Skeletal parameter estimation from optical motion capture data. In *CVPR*, 2005.
- [14] J. F. O’Brien, R. E. Bodenheimer, G. J. Brostow, and J. K. Hodgins. Automatic joint parameter estimation from magnetic motion capture data. In *Graphics Interface*. Canadian Human-Computer Communications Society, 2000.
- [15] M. Paladini, A. D. Bue, J. M. F. Xavier, L. de Agapito, M. Stosic, and M. Dodig. Optimal metric projections for deformable and articulated structure-from-motion. *IJCV*, 2012.
- [16] D. A. Ross, D. Tarlow, and R. S. Zemel. Learning articulated structure and motion. *IJCV*, 2010.
- [17] J. Sturm, C. Stachniss, and W. Burgard. A probabilistic framework for learning kinematic models of articulated objects. *J. Artif. Intell. Res. (JAIR)*, 2011.
- [18] C. Tomasi and T. Kanade. Shape and motion from image streams under orthography: a factorization method. *IJCV*, 1992.
- [19] P. A. Tresadern and I. D. Reid. Articulated structure from motion by factorization. In *CVPR*, 2005.
- [20] R. Tron and R. Vidal. A benchmark for the comparison of 3-d motion segmentation algorithms. In *CVPR*, 2007.
- [21] J. Yan and M. Pollefeys. A factorization-based approach for articulated nonrigid shape, motion and kinematic chain recovery from video. *IEEE TPAMI*, 2008.