Chapter 2: Representing a Moving Scene

Konrad Koniarski

2016/07/12

1 The Orgines of 3D Reconstruction

1.1 The Origins of 3D Reconstruction 1

The goal to reconstruct the three-dimensional structure of the world from a set of two-dimensional view has a long history in computer vision. It is a classical **ill-posed problem**, because the reconstruction consistent with a given set of observations/images is typically not unique. Therefore, one will need to impose additional assumptions. Mathematically, the study of geomeric relations between a 3D scene and the observed 2D projections is based on two types of transformations, namely:

- **Euclidean motion** or **rigid-body motion** representing the motion of the camera from one frame to the next.
- **Perspective projection** to account for the image formation process (see pinhole camera, etc.).

The notion of perspective projection has it roots among the ancient Greeks (Euclid of Alexandria, 400 B.C.) and the Renaissance period (Brunelleschi and Alberti, 1435). The study of perspective projection lead to the field of projective geometry (girard Desargues 1648, Gaspard Monge 18th cent.).

1.2 The Origins of 3D Reconstruction 2

The first work on the problem of multiple view geomery was that of **Erwin Krupp (1913)** who showed that two views of five points are sufficent to determine both the relative transformation (**motion**) between the two views and the 3D location (**structure**) of the points up to finitely many solutions. A linear algorithm to recover structure and motion form two views based on the epipolar constraint was proposed by **Longuet-Higgins in 1981**. An entire series of works along these lines was summarized in several text books (Faugeras 1993, Kanatani 1993, Maybank 1993, Weng et al. 1993). Extensions to three view were developed by Spetsakis and Aloimonos 1987, 1990 and by Shashua 1994 and Hartley 1995. Factorization techniques for multiple view and orthogonal projection were developed by Tomasi and Kanade 1992. The joint estimation of camera motion and 3D location is called **structure and motion** or **visual SLAM**.

2 3D Spaces and Rigid Body Motion

2.1 Three-dimensional Euclidean Space

The three-dimensional Euclidean space $\mathbb{E}^3$ consist of all points $p \in \mathbb{E}^3$ characterized by coordinates

$$X \equiv (X_1, X_2, X_3)^T \in \mathbb{R}^3 \tag{1}$$

such that $\mathbb{E}^3$ can be identified with $\mathbb{R}^3$. That means we talk about points ($\mathbb{E}^3$) and coordinates ($\mathbb{R}^3$) as if they were the same thing. Given to points X nad Y, one can define a **bound vector** as

$$v = Y - X \in \mathbb{R}^3 \tag{2}$$

Consider this vector independent of its base point Y makes it a **free vector**. The set of free vectors $v \in \mathbb{R}^3$ forms a linear vector space. By identifing $\mathbb{E}^3$ and $\mathbb{R}^3$, one can endow $\mathbb{E}^3$ with a scalar product, a norm and a metric. This allows to compute **distances**, **curve length**

$$I(\gamma) \equiv \int_0^1 |\gamma(s)| ds \text{ for a curve : } \gamma : [0,1] \to \mathbb{R}^3 \tag{3}$$

areas or **volumes**.

2.2 Cross Product and Skew-symetric Matrices

On $\mathbb{R}^3$ one can define a cross product

$$\times : \mathbb{R}^3 \times \mathbb{R}^3 \to \mathbb{R}^3 : u \times v = \begin{pmatrix} u_2v_3 - u_3v_2 \\ u_3v_1 - u_1v_3 \\ u_1v_2 - u_2v_1 \end{pmatrix} \in \mathbb{R}^3 \tag{4}$$

which is a **vector orthogonal to u and v**. Since $u \times v = -v \times u$, the cross product introduces an orientation. Fixing u induces a linear mapping $v \to u \times v$ which can be represented by **skew-symetric matrix**

$$\hat{u} = \begin{pmatrix} 0 & -u_3 & u_2 \\ u_3 & 0 & -u_1 \\ -u_2 & u_1 & 0 \end{pmatrix} \in \mathbb{R}^{3\times 3} \tag{5}$$

In turn, every skew-summetric matrix $M = -M^T \in \mathbb{R}^{3\times 3}$ can be identified with a vector $u \in \mathbb{R}^3$. The operator $\hat{}$ defines an **isomorphism** between $\mathbb{R}^3$ and space $so(3)$ of all 3×3 skew-symetric matrices. Its inverse is denoted by $v : so(3) \to \mathbb{R}^3$.

2.3 Rigid-body Motion

A **rigid-body motion** (or rigid-body transformation) is a family of maps

$$g_t : \mathbb{R}^3 \to \mathbb{R}^3, X \to g_t(X), t \in [0, T] \tag{6}$$

which preserve the norm and cross product of any two vectors:

- $|g_t(v)| = |v|, \forall v \in \mathbb{R}^3$,
- $g_t(u) \times g_t(v) = g_t(u \times v), \forall u, v \in \mathbb{R}^3$.

Since norm and scalar product are related by the **polarization identity**

$$\langle u, v \rangle = \frac{1}{4}(|u + v|^2 - |u - v|^2) \tag{7}$$

one can also state that a rigid-body motion is a map which preserves inner product and cross product. As a consequence, rigid-body motions also preserve the triple product

$$\langle g_t(u), g_t(v) \times g_t(w) \rangle = \langle u, v \times w \rangle, \forall u, v, w \in \mathbb{R}^3 \tag{8}$$

which mean that they are volume-preserving.

2.4 Representation of Rigid-body Motion

Since it preserves lengths and orientation, the motion g_t of a rigid body is sufficiently defined by specifying the motion of a Cartesian coordinate frame attached to the object (given by an origin and orthonormal oriented vectors $e_1, e_2, e_3 \in \mathbb{R}^3$). The $T \in \mathbb{R}^3$, whereas the transformation of the vectors e_i is given by new vectors $r_i = g_t(e_i)$.

Scalar and cross product of these vectors are preserved:

$$r_i^T r_j = g_t(e_i)^T g_t(e_j) = e_i^T e_j = \delta_{ij}, r_1 \times r_2 = r_3 \tag{9}$$

The first constraint amounts to the statement that the matrix $R = (r_1, r_2, r_3)$ is an orthogonal (rotation) matrix: $R^T R = RR^T = I$, whereas the second property implies that $det(R) = \pm 1$, in other words: R is an element of the group $SO(3) = \{R \in \mathbb{R}^{3\times 3} | R^T R = I, det(R) = \pm 1\}$. Thus the rigid body motion g_t can be written as:

$$g_t(x) = Rx + T \tag{10}$$

2.5 Exponential Coordinates of Rotation

We will now derive a representation of an **infinitesimal rotation**. To this end, consider a family of rotation matrices R(t) which continuously transform a point from its orginal location (R(0) = I) to a different one.

$$X_{trans}(t) = R(t)X_{orig}, \text{with} R(t) \in SO(3) \tag{11}$$

Since $R(t)R(t)^T = I, \forall t$, we have

$$\frac{d}{dt}(RR^T) = \dot{R}R^T + R\dot{R}^T = 0 \Rightarrow \dot{R}R^T = -(\dot{R}R^T)^T \tag{12}$$

Thus, $\dot{R}R^T$ is a **skew-symmetric matrix**. As shown in the section about the $\hat{\cdot}$-operator, this implies that there exists a vector $w(t) \in \mathbb{R}^3$ such that:

$$\dot{R}(t)R^T(t) = \hat{w(t)} \Leftrightarrow \dot{R}(t) = \hat{w}R(t) \tag{13}$$

Since $R(0) = I$, it follows that $\dot{R}(0) = \hat{w}(0)$. Therefor the **skew-symmetric matrix** $\hat{w}(0) \in so(3)$ **gives the first order approximation of a roation**:

$$R(dt) = R(0) + dR = I + \hat{w}(0)dt \tag{14}$$

3 The Lie Group SO(3)

3.1 Lie Group and Lie Algebra

The above calculations showed that the effect of any infinitesimal rotation $R \in SO(3)$ can be approximated by an element from the space of skew-symmetric matrices

$$so(3) = \{\hat{w} | w \in \mathbb{R}^3\}. \tag{15}$$

The rotation group SO(3) is called a **Lie group**. The space so(3) is called its **Lie algebra**.

Definition: A **Lie group** (or infinitesimal group) is a smooth manifold that is also a group, such that the group operations multiplication and inversion are smooth map.

As show above: **The Lie algebra so(3) is the tangent space as the identity of the rotation group SO(3)**.

An **algebra over a filed K** is a vector space V over K with multiplication on the space V. Elements $\hat{w}$ and $\hat{v}$ of the Lie algebra generally do not commute. One can define the **Lie bracket**

$$[\cdot,\cdot] : so(3) \times so(3) \to so(3); [\hat{w},\hat{v}] \equiv \hat{w}\hat{v} - \hat{v}\hat{w} \tag{16}$$

3.2 The Exponental Map

Given the infinitesimal formulation or rotation in terms of the skew symmetric matrix $\hat{w}$, is it possible to determine a useful representation of the rotation R(t) ? let us assume that $\hat{w}$ is constatnt in time. The differenntial equation system

$$\begin{cases} \dot{R}(t) = \hat{w}R(t) \\ R(0) = I \end{cases} \tag{17}$$

has the solution

$$R(t) = e^{\hat{w}t} = \sum_{n=0}^{\infty} \frac{(\hat{w}t)^n}{n!} = I + \hat{w}t + \frac{(\hat{w}t)^2}{2!} + ... \tag{18}$$

which is a rotation around the axis $w \in \mathbb{R}^3$ by a angle of t (if) $||w|| = 1$). Alternatively, one can absorb the scalar $t \in \mathbb{R}$ into the skew symmetric matrix $\hat{w}$ to obtain $R(t) = e^{\hat{v}}$ with $\hat{v} = \hat{w}t$. This matrix exponential therefore defines a map from the Lie albegra to the Lie group:

$$exp : so(3) \to SO(3); \hat{w} \to e^{\hat{w}} \tag{19}$$

3.3 The Logarithm of SO(3)

As in the case of real analysis one can define an inverse function to the exponential map by the logarithm. In the context of Lie groups, this will lead to a mapping from the Lie group to the Lie algebra. For any rotation matrix $\mathbb{R}^3$ such that $R = exp(\hat{w})$. Such an element is denoted by $\hat{w} = log(R)$ If $R = (r_{ij}) \neq I$, then an appropriate w is given by :

$$|w| = cos^{-1}\{\frac{trace(R) - 1}{2}\}, \frac{w}{|w|} = \frac{1}{2sin(|w|)} \begin{pmatrix} r_{32} - r_{23} \\ r_{13} - r_{31} \\ r_{21} - r_{12} \end{pmatrix} \tag{20}$$

For $R = I$, we have $|w| = 0$, i.e. a rotation by an angle 0. The sbove statement says: Any orthogonal transformation $R \in SO(3)$ can be realized by rotating by an angle $|w|$ around an axis $\frac{w}{|w|}$ as defined above. We will not prove this statement.

Obviously the above representation is not unique since increasing the angle by multiples to 2π will give the same R.

3.4 Rodrigues' Formula

We have seen that any rotation can be realized by computing $R = e^{\hat{w}}$. In analogy to the well-known Euler equation

$$e^{\hat{w}} = I + \frac{\hat{w}}{|w|} sin(|w|) + \frac{\hat{w}^2}{|w|^2}(1 - cos(|w|)) \tag{21}$$

This is known as Rodrigues' formula.

Proof: Let $t = |w|$ and $v = w/|w|$. Then

$$\hat{v}^2 = vv^T - I, \hat{v}^3 = -\hat{v}, ... \tag{22}$$

and

$$\boxed{e^{\hat{w}} = e^{\hat{v}t} = I + \underbrace{(t - \frac{t^3}{3!} + \frac{t^5}{5!} + ...)}_{sin(t)} \hat{v} + \underbrace{(\frac{t^2}{2!} - \frac{t^4}{4!} - \frac{t^6}{6!} - ...)}_{1-cos(t)} \hat{v}^2} \tag{23}$$

4 The Lie Group SE(3)

4.1 Representation of Righd-body Motions SE(3)

We have seen that the motion of a rigid-body is uniquely determined by specifying the translation T oa ny given point and a rotation matrix R derining the transformation of an oriented Cartesian coordinate frame at the given point. Thus the spece of rigid-body motions given by the group of special Euclidean transformation

$$\boxed{SE(3) \equiv g = (R, T) | R \in SO(3), T \in \mathbb{R}^3} \tag{24}$$

In homogeneous coordinates, we have:

$$\boxed{SE(3) \equiv \{g = \begin{pmatrix} R & T \\ 0 & 1 \end{pmatrix} | R \in SO(3), T \in \mathbb{R}^3\} \subset \mathbb{R}^{4\times4}} \tag{25}$$

In the context of rigid motions, one can see the difference between points in $\mathbb{E}^3$ (which can be rotated and translated) and vector in $\mathbb{R}^3$ (which can only be rotated).

4.2 The Lie Algebra o Twists

Given a continous family of rigid-body transromations

$$g : \mathbb{R} \to SE(3); g(t) = \begin{pmatrix} R(t) & T(t) \\ 0 & 1 \end{pmatrix} \in \mathbb{R}^{4\times4} \tag{26}$$

we consider

$$\dot{g}g^{-1}(t) = \begin{pmatrix} \dot{R}R^T & \dot{T} - \dot{R}R^T T \\ 0 & 0 \end{pmatrix} \in 4 \times 4 \tag{27}$$

As in the case of SO(3), the $\dot{R}R^T$ corresponds to some skew-symmetric matrix $\hat{w} \in so(3)$. Defining a vector $v(t) = \dot{(T)}(t) - \hat{(t)}T(t)$, we have:

$$\dot{g}(t)g^{-1}(t) = \begin{pmatrix} \hat{w}(t) & v(t) \\ 0 & 0 \end{pmatrix} \equiv \hat{\xi}(t) \in \mathbb{R}^{4\times4} \tag{28}$$

4.3 The Lie Algebra of Twist

Multiplying with g(t) from the right, we obtain:

$$\dot{g} = \dot{g}g^{-1}g = \hat{\xi}g \tag{29}$$

The 4×4 matrix $\hat{\xi}$ can be viewed as a tangent vector along the curve g(t). $\hat{\xi}$ is called a twist. As in the case of so(3), the set of all twist forms a the tangent space which is the Lie algebra

$$\boxed{se(3) \equiv \{\hat{\xi} = \begin{pmatrix} \hat{w} & v \\ 0 & 0 \end{pmatrix} | \hat{w} \in so(3), v \in \mathbb{R}^3\} \subset \mathbb{R}^{4\times4}} \tag{30}$$

to the Lie group SE(3). As before, we can define operations $\wedge$ and $\vee$ to convert between a twist $\hat{\xi} \in se(3)$ and its twist coordinates $\xi \in \mathbb{R}^6$:

$$\hat{\xi} \equiv \begin{pmatrix} v \\ w \end{pmatrix}^{\wedge} \equiv \begin{pmatrix} \hat{w} & v \\ 0 & 0 \end{pmatrix} \in \mathbb{R}^{4\times4}, \begin{pmatrix} \hat{w} & v \\ 0 & 0 \end{pmatrix}^{\vee} = \begin{pmatrix} v \\ w \end{pmatrix} \in \mathbb{R}^6 \tag{31}$$

4.4 Exponential Coordinates for SE(3)

The twist coordinates $\xi = \begin{matrix} v \\ w \end{matrix}$ are formated by stacking the linear velocity $v \in \mathbb{R}^3$ (related to rotation). The differential equation system

$$\begin{cases} \dot{g}(t) = \hat{\xi}g(t), \hat{\xi} = const. \\ g(0) = I \end{cases} \tag{32}$$

has the solution

$$\boxed{g(t) = e^{\hat{\xi}t} = \sum_{n=0}^{\infty} \frac{(\hat{\xi}t)^n}{n!}} \tag{33}$$

For $w = 0$, we have $e^{\hat{\xi}} = \begin{pmatrix} I & v \\ 0 & 1 \end{pmatrix}$, while for $w \neq 0$ one can show:

$$\boxed{e^{\hat{\xi}} = \begin{pmatrix} e^{\hat{w}} & \frac{(i-e^{hatw})\hat{w}v+ww^Tv}{|w|} \\ 0 & 1 \end{pmatrix}} \tag{34}$$

4.5 Exponential Coordinates for SE(3)

The above shows that the exponential map defines a transformation form the Lie algebra se(3) to the Lie group SE(3):

$$exp : se(3) \to SE(3) : \hat{\xi} \to e^{\hat{\xi}} \tag{35}$$

The elemtns $\hat{\xi} \in se(3)$ are called the **exponential coordinates** for SE(3).

Conversely: **For every** $g \in SE(3)$ **there exist twist coordinates** $\xi = (v, w) \in \mathbb{R}^6$ **such that** $g = exp(\hat{\xi})$

Proof: Given $g = (R, T)$, we know that there exists $w \in \mathbb{R}^3$ with $e^{\hat{w}} = R$. if $|w| \neq 0$, the exponential form of g introduced above shows that we merely need to solve the equation

$$\frac{(I - e^{\hat{w}})\hat{w}v - ww^T v}{|w|} = T \tag{36}$$

for the velocity vector $v \in \mathbb{R}$. Just as in the case of $SO(3)$, this representation is generally not unique, i.e. there exist many twists $\hat{\xi} \in se(3)$ which represent the same rigid-body motion $g \in SE(3)$

5 Representing the Camera Motion

5.1 Representing the Motion of the Camera

When observing a scene from a mobing camera, the coordinates and velocity of points in camera coordinates will change. We will use a rigid-body transformation

$$g(t) = \begin{pmatrix} R(t) & T(t) \\ 0 & 1 \end{pmatrix} \in SE(3) \tag{37}$$

to represent the motion from a fixed world frame to the camera frame at time t. In particular we assume that at time $t = 0$ the camera frame coincides with the world frame, i.e. $g(0) = I$. For any point X_0 in world coordinates, its coordinates in the camera frame at time t are:

$$\boxed{X(t) = R(t)X_0 + T(t)} \tag{38}$$

or in the homogeneous representation

$$X(t) = g(t)X_0 \tag{39}$$

5.2 Concatenated of Motion over Frames

Given two different times t_1 and t_2 we denote the transformation from the points in frame t_1 to the point in frame t_2 by $g(t_2, t_1)$:

$$X(t_2) = g(t_2, t_1)X(t_1) \tag{40}$$

Obviously we have:

$$X(t_3) = g(t_3, t_2)X_2 = g(t_3, t_2)g(t_2, t_1)X(t_1) = g(t_3, t_1)X(t_1) \tag{41}$$

and thus:

$$g(t_3, t_1) = g(t_3, t_2)g(t_2, t_1) \tag{42}$$

By transferring the coordinates of frame t_1 to coordinates in frame t_2 and back, we see that:

$$X(t_1) = g(t_1, t_2)X(t_2) = g(t_1, t_2)g(t_2, t_1)X(t_1) \tag{43}$$

which must hold for any point coordinates $X(t_1)$, thus

$$g(t_1, t_2)g(t_2, t_1) = I \Leftrightarrow g^{-1}(t_2, t_1) = g(t_1, t_2) \tag{44}$$

5.3 Rules of Velocity Transformation

The coordinates of point X_0 in frame t are fiven by $X(t) = g(t)X_0$. Therefore the velocity is given by

$$\dot{X}(t) = \dot{g}(t)X_0 = \dot{g}(t)g^{-1}(t)X(t) \tag{45}$$

By introducing the twist coordinates

$$\hat{V}(t) \equiv \dot{g}(t)g^{-1}(t) = \begin{pmatrix} \hat{w}(t) & v(t) \\ 0 & 0 \end{pmatrix} \in se(3) \tag{46}$$

we got the expression:

$$\boxed{\dot{X}(t) = \hat{V}(t)X(t)} \tag{47}$$

In simple 3D-coordinates this gives:

$$\boxed{\dot{X}(t) = \hat{w}(t)X(t) + v(t)} \tag{48}$$

The symbol $\hat{V}(t)$ therefore represents the ralative velocity of the world frame as viewed from the camera frame.

5.4 Transfer Between Frames: The Adjoint Map

Suppose that w viewer in another frame A is displaced relative to the current frame by a transformation $g_{xy} : Y = g_{xy}X(t)$. Then the velocity in this new frame is given by:

$$\dot{Y}(t) = g_{xy}\dot{X}(t) = g_{xy}\hat{V}(t)X(t) = g_{xy}\hat{V}_{xy}^{-1}Y(t) \tag{49}$$

This shows that the relative velocity of points observed from camera frame A is represented by the twist

$$\hat{V}_y = g_{xy}\hat{V}g_{xy}^{-1} \equiv ad_{g_{xy}}(\hat{V}) \tag{50}$$

where we have introduced the **adjoint map on se(3)**:

$$ad_g : se(3) \to se(3); \hat{\xi} \to g\hat{\xi}g^{-1} \tag{51}$$

	Rotation SO(3)	Rigid-body SE(3)
Matrix representation	$R \in GL(3)$ $R^T R = I$ $det(R) = 1$	$g = \begin{pmatrix} R & T \\ 0 & 0 \end{pmatrix}$
3-D coordinates	$X = RX_0$	$X = RX_0 + T$
Inverse	$R^{-1} = R^T$	$g^{-1} = \begin{pmatrix} R^T & -R^T T \\ 0 & 1 \end{pmatrix}$
Expotential representation	$R = exp(\hat{w})$	$g = exp(\hat{\xi})$
Velocity	$\dot{X} = \hat{w}X$	$\dot{X} = \hat{w}X + v$
Adjoint map	$\hat{w} \to R\hat{w}R^T$	$\hat{\xi} \to g\hat{\xi}g^{-1}$

5.5 Summary

6 Euler Angles

6.1 Alternative Representations: Euler Angles

In addition to the exponential parametrization, there exist alternative mathematical representations to parameterization, there exis alternative mathematical representations to parameterize rotation matrices $R \in SO(3)$, given by the **Euler angles**. These are **local** coordinates, i.e. the parameterization is only correct for a portion of $SO(3)$. Given a basis $(\hat{w}_1, \hat{w}_2, \hat{w}_3)$ of the Lie algebra so(3), we can define a mapping from $\mathbb{R}^3$ to the Lie group SO(3) by

$$\alpha : (\alpha_1, \alpha_2, \alpha_3) \to exp(\alpha_1 \hat{w}_1 + \alpha_3 \hat{w}_3 + \alpha_3 \hat{w}_3). \tag{52}$$

The coordinates $(\alpha_1, \alpha_2, \alpha_3)$ are called **Lie-Cartan coordinates of the first kind** relative to the above basis. The **Lie-Cartan coordinates of the second kind** are defined as

$$\beta : (\beta_1, \beta_2, \beta_3) \to exp(\beta_1 \hat{w}_1) exp(\beta_2 \hat{w}_2) exp(\beta_3 \hat{w}_3) \tag{53}$$

For the basis representing rotation around the z-, y-, x- axis

$$w_1 = (0, 0, 1)^T, w_2 = (0, 1, 0)T, w_3 = (1, 0, 0)^T \tag{54}$$

the coordinates $\beta_1, \beta_2, \beta_3$ are called **Euler angles**.