A Differential Geometric Approach to Multiple View Geometry in Spaces of Constant Curvature

Yi Ma
Electrical & Computer Engineering Department
University of Illinois at Urbana-Champaign
1406 West Green Street, Urbana, IL 61801
Tel: (217)-244-0871   Email: yima@uiuc.edu

December 11, 2001

Abstract. Based upon an axiomatic formulation of vision system in a general Riemannian manifold, this paper provides a unified framework for the study of multiple view geometry in three dimensional spaces of constant curvature, including Euclidean space, spherical space, and hyperbolic space. It is shown that multiple view geometry for Euclidean space can be interpreted as a limit case when (sectional) curvature of a non-Euclidean space approaches to zero. In particular, we show that epipolar constraint in the general case is exactly the same as that known for the Euclidean space but should be interpreted more generally when being applied to triangulation in non-Euclidean spaces. A special triangulation method is hence introduced using trigonometry laws from Absolute Geometry. Based on a common rank condition, we give a complete study of constraints among multiple images as well as relationships among all these constraints. This idealized geometric framework may potentially extend extant multiple view geometry to the study of astronomical imaging where the effect of space curvature is no longer negligible, e.g., the so-called "gravitational lensing" phenomenon, which is currently active study in astronomical physics and cosmology.

Keywords: Multiple view geometry, spaces of constant curvature, gravitational lensing, epipolar constraint, multilinear constraint, algebraic and geometric dependency, triangulation.

1. Introduction

Classic multiple view geometry in a three dimensional *Euclidean* space $\mathbb{E}^3$ ($\mathbb{R}^3$ with its standard inner product as metric) has been extensively studied for the past two decades. A commonly adopted mathematical model for a pin-hole camera can be described as:

$$\lambda(t)\boldsymbol{x}(t) \;=\; APg_e(t)p. \tag{1}$$

In this equation, $A \in SL(3)$ (the group of 3×3 real matrices of determinant 1) is the so-called *calibration matrix*, $P \in \mathbb{R}^{3\times 4}$ is a *standard projection matrix*, and $g_e \in \mathbb{R}^{4\times 4}$ is a homogeneous representation for a *Euclidean motion* (denoted by $SE(3)$) with a rotation $R(t) \in SO(3)$ (the group of 3×3 rotation matrices of determinant 1) and a translation $T(t) \in \mathbb{R}^3$ w.r.t. the coordinate frame at time $t = 0$:

$$P = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \in \mathbb{R}^{3\times 4}, \qquad g_e(t) = \begin{bmatrix} R(t) & T(t) \\ 0 & 1 \end{bmatrix} \in \mathbb{R}^{4\times 4}. \quad (2)$$

Then $\boldsymbol{x}(t) = [x_1(t), x_2(t), x_3(t)]^T \in \mathbb{R}^3$ is the *image* at time t (in homogeneous coordinates) of a *point* $p = [p_1, p_2, p_3, 1]^T \in \mathbb{R}^4$ (also in homogeneous coordinates). Apparently, $\boldsymbol{x}$ differs from APg_ep by an arbitrary positive scalar $\lambda(t) \in \mathbb{R}_+$. In the literature, sometimes the above equation is written as:

$$\lambda \boldsymbol{x} = [AR, AT]p.$$

The fundamental problem for multiple view geometry is then to study how to recover from multiple images $\{\boldsymbol{x}(t)\}$ of a set of points: the camera motion $g_e(t)$ and the 3-D coordinates of the points $\{p\}$.

A widely adopted approach to study this problem is a three-stage stratification through a *projective* and then an *affine* reconstruction (Faugeras, 1995). That is, instead of directly dealing with the Euclidean motion group $SE(3)$, one introduces two intermediate cases where $SE(3)$ is modified by a *projective transformation* g_p in the group $GL(4, \mathbb{R})$ or an *affine transformation* g_a in $A(3, \mathbb{R})$.[1] Respectively, the camera model (1) becomes:

$$\begin{aligned} & \lambda(t)\boldsymbol{x}(t) = APg_e(t)g_pp, \quad g_p \in GL(4, \mathbb{R}), \\ \text{or} \quad & \lambda(t)\boldsymbol{x}(t) = APg_e(t)g_ap, \quad g_a \in A(3, \mathbb{R}). \end{aligned} \qquad (3)$$

Multiple view geometry for these two camera models has been well established in the literature (see (Faugeras, 1995; Hartley and Zisserman, 2000) for details). Exercising the same philosophy, we here may further consider a more generalized camera model:

$$\lambda(t)\boldsymbol{x}(t) = APg(t)p, \qquad (4)$$

where $g(t)$ belongs to G which could be any (Lie) subgroup embedded in $GL(4, \mathbb{R})$. The reader should be aware that there is a difference between this model (4) and the above stratification model (3). For (4), the transformation from one camera frame to another can be an arbitrary element in G. But for (3) there is essentially only *one* global projective or affine transformation which acts on all camera coordinate frames simultaneously. Between camera frames the transformation is typically *not* a free projective or affine transformation.

Certainly, the Euclidean group $SE(3)$ and affine group $A(3, \mathbb{R})$ are two candidates for such G. However, as we will soon see, there are other subgroups embedded in $GL(4, \mathbb{R})$ which are not necessarily subgroups of $A(3, \mathbb{R})$ nor $SE(3)$. It is then important to quest:

(i) *Whether these groups (other than $SE(3)$) also represent meaningful vision problems? Are these problems of any practical applications?*

On the other hand, conceptually there is obvious limitation for all three groups $SE(3), A(3,\mathbb{R})$ and $GL(4,\mathbb{R})$: they only allow us to study vision in linear spaces (Euclidean $\mathbb{R}^3$ or Projective $\mathbb{P}^3$) in which light rays travel straight lines. Hence, multiple view geometry developed so far is based on a default assumption: the underlying space is essentially a Euclidean space $\mathbb{E}^3$. Mathematically, it is then natural to ask:

(ii) *If the Euclidean assumption on the underlying space is removed, can we still study vision or multiple view geometry, and how?*

In order to answer this question, we need clearly understand what are *all* the hidden assumptions which have essentially enabled the development of multiple view geometry so far and how these assumptions can be re-formulated in a more general form so as to accommodate cases of non-Euclidean spaces.

In this paper, we attempt to provide a definite answer to questions **(i)** and **(ii)**, which are two questions in fact deeply related. Basically, we will show that, under certain assumptions, it is possible to generalize multiple view geometry to non-Euclidean spaces by choosing for G in the model (4) subgroups of $GL(4,\mathbb{R})$ other than $SE(3)$ and $A(3,\mathbb{R})$. From this group theoretic viewpoint, most results that we have obtained for the Euclidean space have their natural extensions to the non-Euclidean case. The Euclidean case in principle can be interpreted as an extreme case of the non-Euclidean multiple view geometry. We hope that such a generalization not only captures essential geometric characteristics of any imaging system but also provides a unified mathematical framework in which we may gain a deeper understanding of underlying principles of multiple view geometry in general.

This paper aims to provide a theoretic treatment of multiple view geometry from a differential geometric point of view. Since the geometry of non-Euclidean spaces is a well established subject (Kobayashi and Nomizu, 1996; Wolf, 1984), we here try not to over address it and only to take whatever results available to serve our own purpose. Although main techniques in this paper involve only linear algebra and matrix (Lie) groups, a background in differential geometry (especially in Riemannian geometry and Lie groups) will certainly improve your appreciation of some general concepts and special subjects to be introduced below. Now that we are developing a multiple view geometry for non-Euclidean spaces in parallel to that known for Euclidean space, references for relevant facts for the Euclidean case will be given along the development. In case we miss, we point the reader to the book (Hartley and Zisserman, 2000), which gives a rather complete and comprehensive summary of Euclidean multiple view geometry.

2. Vision and Imaging in Riemannian Manifolds

Not until Einstein's theory of general relativity, non-Euclidean geometry, or Riemannian geometry in general, is more of a pure mathematical creation rather than a geometry of physical meaning. According to general relativity, any physical space is in fact described by a Riemannian manifold and its curvature is attributed to the distribution of mass in the space. In such a space, light travels the so-called *geodesic*, i.e. the curve of minimal distance among all those connecting two given points. Such a curve is in general "bent" by the field of gravity. If the density of mass is small, the space has almost zero curvature hence can be approximately assumed to be a Euclidean space $\mathbb{E}^3$. Geodesics in this space are then nothing but the straight lines.

2.1. Gravitational lensing

The effect of non-zero space curvature on astronomical imaging has only recently been demonstrated by modern deep space telescopes, e.g., the Hubble telescope.[2] Below is an example image released by NASA:[3]

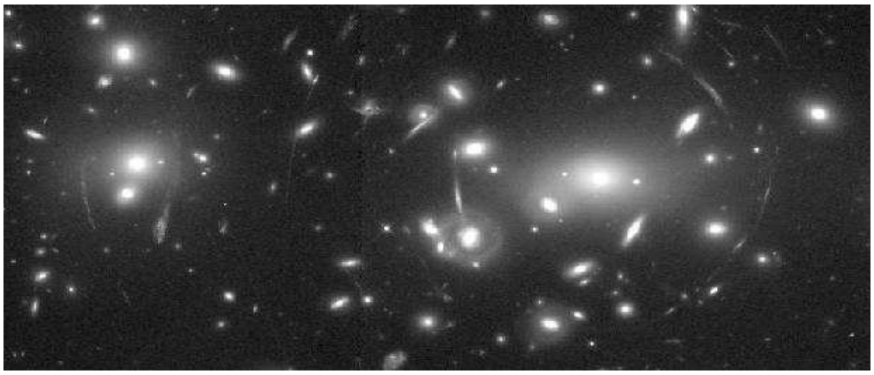

Figure 1. Explanation: Gravity can bend light. Almost all of the bright objects in this Hubble Space Telescope image are galaxies in the cluster known as Abell 2218. The cluster is so massive and so compact that its gravity bends and focuses the light from galaxies that lie behind it. As a result, multiple images of these background galaxies are distorted into faint stretched out arcs - a simple lensing effect analogous to viewing distant street lamps through a glass of wine. The Abell 2218 cluster itself is about 3 billion light-years away in the northern constellation Draco.

This phenomenon is referred to as "gravitational lensing" in astronomical physics or cosmology. Illustratively, it can be explained by Figure 2.[4] There are a vast body of literature and images on gravitational lenses, for example see (Schneider, Ehlers and Falco, 1992).

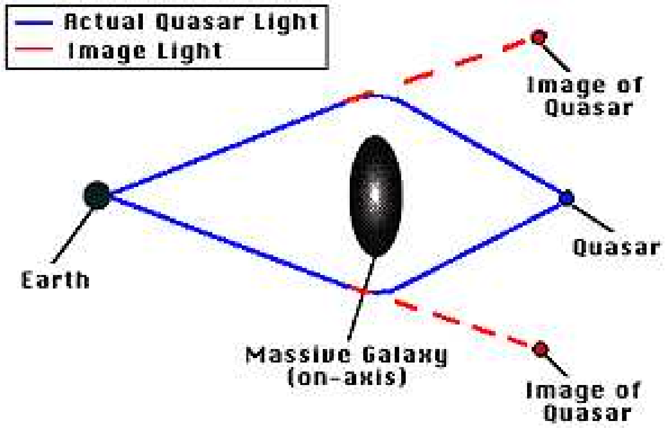

Figure 2. Albert Einstein predicted that the gravitational field of a massive galaxy would bend light traveling to Earth from distant quasars. This is what is called "gravitational lensing", since the intervening galaxy acts as a lens to focus the image of the distant quasar to a new location. Gravitational lensing can produce *multiple* images, rings, or arcs, depending on the distribution of mass in the galaxy and the Earth-galaxy-quasar geometry.

2.2. An Axiomatic Formulation

In this paper, instead of studying the physical nature of gravitational lensing in real-world scenarios, we will however consider a more idealistic situation. The idealization will essentially allow us to gain a clear understanding of geometric laws which govern imaging and vision in any space of non-zero curvature. Now, out of curiosity, let us first imagine an intelligent creature living *in* a sphere $\mathbb{S}^2$ as illustrated in Figure 3 – the simplest ideal example of a 2 dimensional non-Euclidean space with a constant curvature. Unlike in a Euclidean space, light now travels great circles of the sphere instead of straight lines. Then what kind of multiple view geometry the creature would have developed? To answer this question, we need to put ourselves in the shoes of this creature and try to understand what are the basic elements that a vision system in such a space must consist of. In this section, we give an *axiomatic* formulation of these basic elements. Although the proposed mathematical model seems to be given in a rather abstract manner, it is indeed a natural generalization of the conventional camera model in the Euclidean space $\mathbb{E}^3$. Such a generalization allows us to fully discover the geometric nature of any vision system in a very concise and precise way, as we will see in later sections.

Let us consider a (connected) Riemannian manifold (M, Φ), i.e. a differentiable manifold equipped with a positive definite symmetric 2-form Φ as its metric. If the reader is not familiar with differential geometry, he or she may simply replace (M, Φ) by the Euclidean space $\mathbb{E}^3$ with its standard inner product metric. In this paper, we will be mostly interested in three

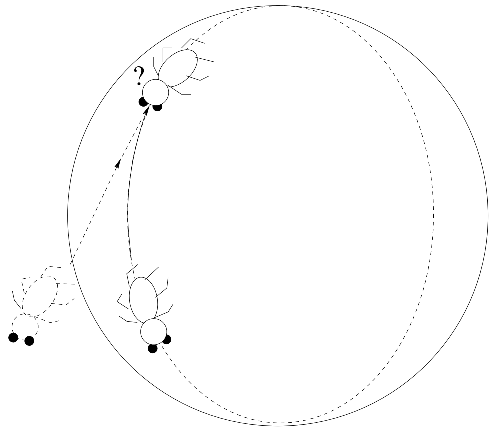

Figure 3. Two 2D bugs live in a 2D sphere. How could the bug tell from what it sees the other bug's correct position and motion *in* this sphere? Certainly it must be aware that the space is not flat and light is bent. Otherwise, it would think the other bug were at somewhere outside of the sphere.

dimensional spaces although the model given below is for the most general case.

Assumption 1 (Camera). *A camera is modeled as a point $o \in M$, which usually stands for the optical center of the camera, and an orthonormal coordinate chart is chosen on T_oM, the tangent space of M at the point o.*

Assumption 2 (Motion). *M is a complete and orientable Riemannian manifold. G is an orientation preserving subgroup of the isometry group of M. The group G then models valid motions of the camera. Its representation might depend on the position of the optical center o.*

Assumption 3 (Light). *In the manifold M, light always travels along the geodesics with a constant speed. For simplicity, here we may assume this speed to be infinite.*

Assumption 4 (Image). *The image of a point $p \in M$ is the ray in the tangent space T_oM which corresponds to the direction of the geodesic connecting p and the optical center o.*

Assumption 5 (Calibration). *The effect of camera calibration can be modeled as an unknown linear transformation $\psi : T_oM \to T_oM$ (as a vector space). In the calibrated case, one may assume this transformation is known or simply the identity map.*

Assumptions 1 to 5 formally define a vision system (hence the camera model) in a class of Riemannian manifolds. When the manifold M happens to be the Euclidean space $\mathbb{E}^3$, the so obtained model is exactly equivalent to the conventional model (1) that we have introduced above. Even in the most general case, the model is based on direct geometric intuition. The only difference is that the manifold M (representing the world space) is explicitly distinguished from the image space T_oM. The reason is that if the scale of the viewer (or the camera) is significantly smaller compared to that of the manifold, the space appears to the viewer as (locally) Euclidean and the tangent space at the standpoint of the viewer is a best approximation for that purpose. So it is only the direction in which the light is from can be detected by the viewer (although the viewer may have the knowledge that the space is globally not Euclidean). In the Euclidean case, the manifold and its tangent space happen to coincide. Intuitively, this general model of vision can be illustrated in the Figure 4.

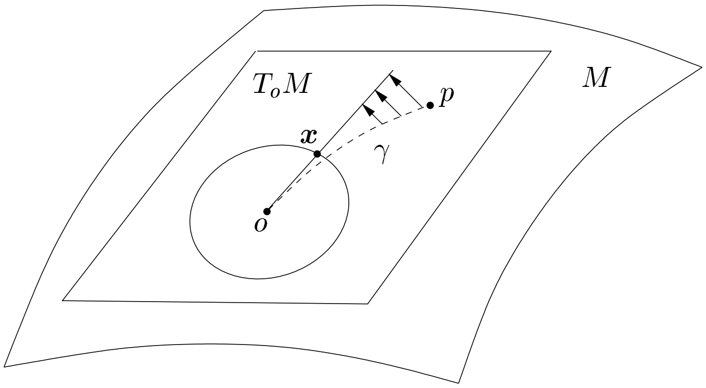

Figure 4. The curve γ is the geodesic connecting camera center o and a point $p \in M$; arrows mean the inverse of the *exponential map* $\exp : T_oM \to M$; $\boldsymbol{x}$ then represents the image of the point p with respect to a camera centered at the point o.

Comment 1 (A Lie Group Viewpoint). *The Lie group G which models the motion of the camera is obtained in the model as being a subgroup of the isometry group of M. In fact the relation between G and M is symmetric at least in the case that the motion group G acts transitively on M: letting G_H be the isotropy subgroup*[5] *of G, then the manifold M is simply the quotient space G/G_H. The Riemannian metric Φ on M can be induced from canonical metrics of G and G_H by this quotient. In practice, this viewpoint is far more useful than the above axiomatic definition since, as we will soon see, most manifolds that we are interested are usually given as submanifolds of a Euclidean space which are invariant under the action of certain Lie groups G. Therefore, geometric property of a vision system in such manifolds is intrinsically inherited from that of G.*

Comment 2 (Classification of Spaces). *As pointed out by Weinstein (*Weinstein, 2000*), different requirements on the properties of the motion group*

G may determine the type of manifold that M must be. For example, if we require G act transitively on the frame bundle[6] *of M, it can be shown that M must be* spaces of constant curvature *(*Kobayashi and Nomizu, 1996*). A less restrictive requirement on G is to allow that the optical axis of the camera can point to any direction at any point of M (but the camera may not be able to rotate arbitrarily around the axis). In this case, M corresponds to the so-call* symmetric spaces of rank 1. *One can further relax the Assumption 2 so that G does not have to be a subgroup of the isometry group of M. Then M can be any Riemannian manifold. Little is known how to study geometric properties of symmetric spaces from a vision point of view much less for Riemannian manifolds in general. Vision theory for general Riemannian manifolds is out of the scope of this paper. For the remaining of this paper, we will focus only on the spaces of constant curvature and demonstrate how to generalize the extant multiple view geometry for Euclidean space to those non-Euclidean spaces.*

3. Multiple View Geometry in Spaces of Constant Curvature

Can the abstract model introduced in the preceding section be of any use? In this section we will demonstrate that, using this model, one can actually extend extant multiple view geometry developed for the Euclidean space to a larger class of spaces: the *spaces of constant curvature*. In particular, we will show that the epipolar geometry has no peculiar meaning to Euclidean space. It is also true for more general spaces. In a similar fashion, dependency among multilinear constraints can be uniformly established for the entire class of spaces of constant curvature.

Spaces of constant curvature are Riemannian manifolds with constant sectional curvature. In differential geometry, they are also referred to as basic *space forms*. A Riemannian manifold of constant curvature is said to be *spherical*, *hyperbolic*, or *flat* (or *locally Euclidean*) as the sectional curvature is positive, negative, or zero, respectively. See Figure 5. Geometry

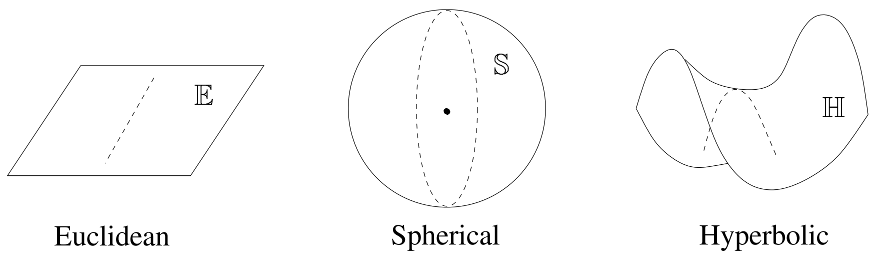

Figure 5. Three basic space forms: Euclidean, spherical and hyperbolic. Conventionally only multiple view geometry in the first type of spaces has been studied.

about spaces of constant curvature is also called *absolute geometry*, named by one of the co-founders of non-Euclidean geometry: Janos Bolyai (Hsiang, 1995). For the rest of this paper, we will try to develop a vision theory for these spaces, as a natural generalization of the extant vision theory for 3 dimensional (3-D) Euclidean space. More specifically,

> *We try to identify all the geometric relationships (or laws) that govern multiple images of an object in a space of constant curvature, from which a reconstruction of the object and camera locations can be achieved.*

We will focus on 3 dimensional spherical and hyperbolic spaces since the Euclidean case has been well understood. On the other hand, as shown in Appendix A, the Euclidean case can always be viewed as a limit case of the general theory.

3.1. Homogeneous Representation of Spaces of Constant Curvature

Geometric properties of n dimensional spaces of constant curvatures have been well studied in differential geometry (as an important case of *symmetric spaces*), e.g., see (Kobayashi and Nomizu, 1996). In the rest of this section, we briefly *review* some of the main results which serve for our purposes. The main goal is to establish necessary mathematical basis for the derivation of the projection model (23) in the next section. However, if the reader is not familiar with differential geometry or Lie groups, this section (i.e. Section 3.1) can be skipped at first read as long as the reader is willing to take for granted the three propositions developed in this section and later the camera model (23) as a result of them.

The first proposition which characterizes the 3 dimensional space of constant curvatures follows directly from a more general statement about n dimensional spaces, see Theorem 3.1 in Chapter V of (Kobayashi and Nomizu, 1996) :

Proposition 1 (Three Dimensional Spaces of Constant Curvature). *Let* $[x_1, x_2, x_3, x_4]^T$ *be the coordinate system of* $\mathbb{R}^4$ *and* M *be the hyper-surface of* $\mathbb{R}^4$ *defined by:*

$$x_1^2 + x_2^2 + x_3^2 + r x_4^2 = r \quad (r\text{: a nonzero constant}). \tag{5}$$

Let g *be the Riemannian metric of* M *obtained by restricting the following second order differential form to* M*:* $dx_1^2 + dx_2^2 + dx_3^2 + r\, dx_4^2$*. Then:*

1. M *is a 3 dimensional space of constant curvature with sectional curvature* $1/r$.
2. *The group* G *of linear transformations of* $\mathbb{R}^4$ *leaving the quadratic form* $x_1^2 + x_2^2 + x_3^2 + r x_4^2$ *invariant acts transitively on* M *as the group of isometry of* M.

3. *If $r > 0$, then M is isometric to a sphere of a radius $\sqrt{r}$. If $r < 0$, then M consists of two mutually isometric connected hyperbolic surfaces in $\mathbb{R}^4$, each of which is diffeomorphic to $\mathbb{R}^3$.*

Now let Q be the 4×4 matrix associated to the quadratic form defining M: $Q = \begin{bmatrix} I_3 & 0 \\ 0 & r \end{bmatrix}$. The *isometry group* G of M is given as a subgroup of $GL(4, \mathbb{R})$ that preserves this quadratic form:

$$G = \left\{ g \in \mathbb{R}^{4\times 4} \mid g^T Q g = Q \right\}.$$

Any element $g \in G$ then has the form:

$$g = \begin{bmatrix} W & y \\ z^T & w \end{bmatrix} \quad \in \mathbb{R}^{4\times 4} \tag{6}$$

with $W \in \mathbb{R}^{3\times 3}, y \in \mathbb{R}^3, z \in \mathbb{R}^3, w \in \mathbb{R}$ and the conditions:

$$\boxed{W^T W + r \cdot z z^T = I_3, \quad W^T y + r \cdot w z = 0, \quad y^T y + r \cdot w^2 = r.} \tag{7}$$

The *isotropy group* G_H of G which leaves the origin of a coordinate frame (or usually the center of the camera frame) $o = [0, 0, 0, 1]^T \in M$ fixed is isomorphic to the orthogonal group $O(3)$:[7]

$$G_H = \left\{ \begin{bmatrix} R & 0 \\ 0 & 1 \end{bmatrix} \in \mathbb{R}^{4\times 4} \mid R \in O(3) \right\}. \tag{8}$$

As we have discussed before, the manifold M can then be identified with the *quotient space* G/G_H.[8]

It follows that the *Lie algebra*[9] $\mathfrak{g}$ of the group G (as a Lie group) is the set of the matrices of the form:

$$\xi = \begin{bmatrix} a & b \\ c^T & 0 \end{bmatrix} \quad \in \mathbb{R}^{4\times 4} \tag{9}$$

where $a \in \mathbb{R}^{3\times 3}, b \in \mathbb{R}^3$ and $c \in \mathbb{R}^3$ satisfy the conditions:

$$\boxed{a^T + a = 0, \quad b + rc = 0.} \tag{10}$$

Let $\mathfrak{m}$ be the linear subspace of the Lie algebra $\mathfrak{g}$ of G consisting of matrices of the form: $\begin{bmatrix} 0 & b \\ c^T & 0 \end{bmatrix} \in \mathbb{R}^{4\times 4}$ with $b, c \in \mathbb{R}^3$ and $b + rc = 0$. Let $\mathfrak{h}$ be the Lie algebra of G_H as a subspace of $\mathfrak{g}$ which consists of matrices of the form: $\begin{bmatrix} a & 0 \\ 0 & 0 \end{bmatrix} \in \mathbb{R}^{4\times 4}$ with $a \in \mathbb{R}^{3\times 3}$ and $a^T + a = 0$. Then we have a *canonical decomposition*:

$$\mathfrak{g} = \mathfrak{h} + \mathfrak{m}.$$

It is direct to check the following relations between the subspaces hold:

$$[\mathfrak{h}, \mathfrak{h}] \subset \mathfrak{h}, \quad [\mathfrak{h}, \mathfrak{m}] \subset \mathfrak{m}, \quad [\mathfrak{m}, \mathfrak{m}] \subset \mathfrak{h} \tag{11}$$

where $[\cdot, \cdot]$ stands for *Lie bracket*.[10] Let $\mathfrak{h}$ be the *vertical tangent subspace* of G and $\mathfrak{m}$ be the *horizontal tangent subspace*. Then according to Theorem 11.1 in Chapter II of Kobayashi (Kobayashi and Nomizu, 1996), this decomposition gives a *canonical connection* on the principle bundle $G(G/G_H, G_H)$ which in turn induces constant sectional curvature $1/r$ on $G/G_H = M$.

The canonical decomposition of the Lie algebra $\mathfrak{g}$ of G results in a decomposition of the group G into two basic actions. One is the rotation around a fixed point, characterized by the subgroup G_H or the subalgebra $\mathfrak{h}$; the other is the translation along geodesics of the manifold M which is obviously related to the subalgebra $\mathfrak{m}$. Denote the quotient map from G to G/G_H as $\pi : G \to G/G_H$. Let $\exp(\cdot)$ be the exponential map from $\mathfrak{g}$ to G.[11] Then according to Theorem 3.2 in Chapter XI of (Kobayashi and Nomizu, 1996), we have:

Proposition 2 (Characterization of Geodesics). *Consider the three dimensional space of constant curvature $M(= G/G_H)$ as above. For each $X \in \mathfrak{m}$, $\pi(\exp(tX)) = \exp(tX) \cdot o$ is a geodesic starting from o and, conversely, every geodesic from o is of this form.*

Now let G_T be the subset of G consisting of all the matrices of the form $\exp(X)$ with $X \in \mathfrak{m}$:

$$G_T = \left\{ \exp(X) \in \mathbb{R}^{4\times 4} \mid X \in \mathfrak{m} \right\}. \tag{12}$$

Then G_T corresponds to the so-called *transvection* on M, an analogy to *translation* in the Euclidean space. Notice that in general G_T is not a subgroup of G (although it is in the Euclidean case) since its representation depends on the base point o. On the other hand, the isotropy subgroup G_H of G corresponds to *rotation* on M. As in the Euclidean case, for a "rigid body motion" on M, it is natural to consider the rotation is in the special orthogonal group $SO(3)$ instead of the full orthogonal group $O(3)$. One of the reasons for only considering $SO(3)$ is that it preserves the orientation of the space.

First notice that, as in the Euclidean case, the transvection set G_T of the isometry group G acts transitively on a space M of constant curvature.[12] Then for any $g \in G$, there exists $g_t \in G_T$ such that $g_t^{-1}(g(o)) = o$, i.e. $g_t^{-1}g$ fixes the origin o. So $g_t^{-1}g \in G_H$, the isotropy group of G. We call this element $g_h = g_t^{-1}g$. It then follows that the group G is equal to:

$$G = G_T G_H. \tag{13}$$

This is the so-called *Cartan decomposition*. Hence for any motion $g \in G$ in the space M, it can always be written as the (matrix) product of a transvection

g_t and a rotation g_h:

$$g = g_t g_h \tag{14}$$

where g_h is of the form (8). We now determine the general expression for $g_t = \exp(X)$.

According to Proposition 2, any geodesic connecting a point $p = [p_1, p_2, p_3, p_4]^T \in M$ to the origin o has the form: $p = \exp(X) \cdot o$ for some $X \in \mathfrak{m}$. Without loss of generality, we may assume that X has the form:

$$X = \begin{bmatrix} 0 & b \\ -b^T/r & 0 \end{bmatrix} \quad \in \mathbb{R}^{4\times 4} \tag{15}$$

for some vector $b \in \mathbb{R}^3$. To simply the notation, define $\eta = \|b\|$ and a vector $T = b/\eta \in \mathbb{R}^3$ of unit length. Here, we use the notation T to represent a translational (or more precisely, transvectional) vector. We may extend the functions $\sin(\cdot)$ and $\cos(\cdot)$ to analytic functions defined on the entire complex plane $\mathbb{C}$:

$$\sin(u) = \frac{1}{2i}(e^{iu} - e^{-iu}), \quad \cos(u) = \frac{1}{2}(e^{iu} + e^{-iu}), \quad u \in \mathbb{C}. \tag{16}$$

Also define $\rho = \sqrt{1/r} \in \mathbb{C}$. Then through direct calculation of the exponential of the matrix X, we get:

$$\exp(X) = \begin{bmatrix} W & y \\ z^T & w \end{bmatrix} = \begin{bmatrix} I_3 + (\cos(\eta\rho) - 1)TT^T & \rho^{-1}\sin(\eta\rho)T \\ \rho\sin(\eta\rho)T^T & \cos(\eta\rho) \end{bmatrix}. \tag{17}$$

Proposition 3 (Rigid Body Motion in Spaces of Constant Curvature). *We can always decompose a general rigid body motion $g \in G$ in a 3 dimensional space of constant curvature M into the multiplication of a transvection g_t and a rotation g_h as*

$$\boxed{g = g_t g_h = \begin{bmatrix} I_3 + (\cos(\eta\rho) - 1)TT^T & \rho^{-1}\sin(\eta\rho)T \\ \rho\sin(\eta\rho)T^T & \cos(\eta\rho) \end{bmatrix} \begin{bmatrix} R & 0 \\ 0 & 1 \end{bmatrix}} \tag{18}$$

where $R \in SO(3)$, $T \in \mathbb{S}^2$, and $\eta \in \mathbb{R}_+$. R is the rotation. The unit vector T is the direction and η is the distance of the translation.

3.2. Projection Model in Spaces of Constant Curvature

Based upon the results given in the previous section, we are now ready to study multiple view geometry in the spaces of constant curvature. Similar to the Euclidean case, we first need to specify the (valid) motion of the camera and the projection model of the camera, i.e. how a 2 dimensional image is formed in spaces of constant curvature. Basically, we will formally establish the fact that the camera model for a three dimensional space M of constant curvature does fit in the general camera model (4) introduced before:

$$\boxed{\lambda(t)\boldsymbol{x}(t) = APg(t)p,} \tag{19}$$

where $g(t)$ is a rigid body motion on M as explicitly represented by Proposition 3.

A point p, in a space M of constant curvature, can be represented in *homogeneous coordinates* as $p = [p_1, p_2, p_3, p_4]^T \in \mathbb{R}^4$ which satisfies the quadratic form: $p_1^2 + p_2^2 + p_3^2 + rp_4^2 = r$ with $1/r$ the sectional curvature of M. Then under the motion $g(t) \in G,\ t \in [t_0, t_f] \subset \mathbb{R}$ of the camera, the homogeneous coordinates of the point p (with respect to the camera frame) satisfy the transformation:

$$p(t) = g(t)p, \quad p(t_0) = p. \tag{20}$$

Notice that, with this representation, the point $o = [0, 0, 0, 1]^T \in \mathbb{R}^4$ is always in M. We then call the point o the *origin* of M. Without loss of generality, this origin is always identified with the center of the camera. So when the origin moves, the coordinates of any point p in M change according to (20).

Now consider the geodesic connecting the origin $o = [0, 0, 0, 1]^T$ to p. According to Propositions 2 and 3, we have:

$$\begin{aligned} p \;=\; \exp(X) \cdot o &= \begin{bmatrix} I_3 + (\cos(\eta\rho) - 1)TT^T & \rho^{-1}\sin(\eta\rho)T \\ \rho\sin(\eta\rho)T^T & \cos(\eta\rho) \end{bmatrix} \begin{bmatrix} 0 \\ 1 \end{bmatrix} \\ &= \begin{bmatrix} \rho^{-1}\sin(\eta\rho)T \\ \cos(\eta\rho) \end{bmatrix} \quad \in \mathbb{R}^4. \end{aligned} \tag{21}$$

Therefore, the unit vector T is equal to:

$$T = \frac{[p_1, p_2, p_3]^T}{\sqrt{p_1^2 + p_2^2 + p_3^2}} \quad \in \mathbb{S}^2.$$

This is exactly the unit tangent vector of M at the origin o pointing in the direction of the point p. Or in other words, the geodesic connecting the origin to the point $p = [p_1, p_2, p_3, p_4]^T \in M$ has its tangent vector at the origin as $T \in T_o(M)$ given in the above.

Combining Assumption 4 and the above discussion, the image of a point p at time t can then be any vector $\boldsymbol{x} \in \mathbb{R}^3$ on the ray $\{\mu T \mid \mu \in \mathbb{R}_+\}$. Therefore, in homogeneous coordinates, the image $\boldsymbol{x}$ of the point p differs from the vector $[p_1(t), p_2(t), p_3(t)]^T$ by an unknown scalar $\lambda \in \mathbb{R}_+$. Still define the *standard projection matrix* $P = [I, 0] \in \mathbb{R}^{3\times 4}$ as before. We then have the relationship:

$$\lambda(t)\boldsymbol{x}(t) = Pp(t) = Pg(t)p. \tag{22}$$

We call the scalar $\lambda(t)$ the *scale* of the point p with respect to the image $\boldsymbol{x}(t)$ at time t. λ is different from η which is exactly the geodesic distance from o to p. They are related by

$$\lambda = \rho^{-1}\sin(\eta\rho).$$

Then both the scale λ and η encode the depth information of the point p. Furthermore, if the calibration of the camera is unknown, the transformation $\psi : T_oM \to T_oM$ in Assumption 5 can be represented by a non-singular 3×3 matrix A since T_oM is isomorphic to $\mathbb{R}^3$ as a vector space. Then the above model for the camera is modified to:

$$\lambda(t)\boldsymbol{x}(t) = APg(t)p \tag{23}$$

where $g(t) \in G \subset \mathbb{R}^{4\times 4}$ is a rigid body motion on M (as we have claimed in the beginning of this section).

3.3. Geometry and Reconstruction from Pairwise Views

In this section, we study the relationship among two images of a point $p \in M$ subject to a rigid body motion of a camera in a space M of constant curvature. According to the Cartan decomposition (Proposition 3), we know that a rigid body motion of the camera can always be expressed in the form: $g = g_t \cdot g_h,\ g_t \in G_T,\ g_h \in G_H$. The transvection part g_t and rotation part g_h respectively have the forms:

$$g_t = \exp(X) = \begin{bmatrix} W & y \\ z^T & w \end{bmatrix}, \quad g_h = \begin{bmatrix} R & 0 \\ 0 & 1 \end{bmatrix}, \tag{24}$$

where $X \in \mathfrak{m},\ R \in SO(3)$ and expressions for $W \in \mathbb{R}^{3\times 3}, y \in \mathbb{R}^3, z \in \mathbb{R}^3$ and $w \in \mathbb{R}$ are given by (17). Then the projection model (23) becomes

$$\lambda(t)\boldsymbol{x}(t) = [AW(t)R(t), Ay(t)]p. \tag{25}$$

3.3.1. *Epipolar Constraints in Spaces of Constant Curvature*

We first assume that the camera is calibrated, i.e. $A = I_3$. Denote the images of $p = [p_1, p_2, p_3, p_4]^T$ before and after a rigid body motion g as $\boldsymbol{x}_1 \in \mathbb{R}^3$ and $\boldsymbol{x}_2 \in \mathbb{R}^3$, respectively. Then according to (23) or the equation above we have: $\lambda_1\boldsymbol{x}_1 = Pp$ and $\lambda_2\boldsymbol{x}_2 = Pgp$. They yield:

$$\begin{aligned} &\lambda_2\boldsymbol{x}_2 = WR\cdot\lambda_1\boldsymbol{x}_1 + p_4 y \ \Rightarrow\ y \times \lambda_2\boldsymbol{x}_2 = y \times (WR\cdot\lambda_1\boldsymbol{x}_1) \\ \Rightarrow\ &\boldsymbol{x}_2^T\widehat{y}WR\boldsymbol{x}_1 = 0. \end{aligned} \tag{26}$$

We in this paper use $\widehat{u}$ to denote the skew symmetric matrix associated to a vector $u \in \mathbb{R}^3$ such that $\widehat{u}v = u \times v$ for any $v \in \mathbb{R}^3$.

In the Euclidean case, (26) would exactly give the well-known bilinear epipolar constraint. As noticed, in the general case, the role of essential matrix is replaced by $\widehat{y}WR$. We need to study the structure of this matrix. According to (17), we have:

$$W = I_3 + (\cos(\eta\rho) - 1)TT^T, \quad y = \rho^{-1}\sin(\eta\rho)T. \tag{27}$$

Notice that we always have $\widehat{T}TT^T = 0$. Suppose $\sin(\eta\rho) \neq 0$. Then (26) yields:

$$\begin{aligned} &\boldsymbol{x}_2^T\widehat{T}WR\boldsymbol{x}_1 = 0 \;\Leftrightarrow\; \boldsymbol{x}_2^T\widehat{T}(I_3 + (\cos(\eta\rho) - 1)TT^T)R\boldsymbol{x}_1 \\ \Leftrightarrow\;& \boldsymbol{x}_2^T\widehat{T}R\boldsymbol{x}_1 = 0. \end{aligned} \tag{28}$$

This is exactly the well-known bilinear *epipolar constraint*. Here we see that this constraint holds for all spaces of constant curvature. The same as in the Euclidean case, we call $E = \widehat{T}R$ the *essential matrix*. As a summary of the above discussion, we have the following theorem:

Theorem 1 (Calibrated Epipolar Constraint). *Consider a rigid body motion of a camera in a space M of constant curvature. If $T \in \mathbb{R}^3$ is the vector associated to the direction of the translation and $R \in SO(3)$ the rotation, then the images $\boldsymbol{x}_1 \in \mathbb{R}^3$ and $\boldsymbol{x}_2 \in \mathbb{R}^3$ of a point $p \in M$ before and after the motion satisfy the epipolar constraint:*

$$\boxed{\boldsymbol{x}_2^T\widehat{T}R\boldsymbol{x}_1 = 0.} \tag{29}$$

Corollary 1 (Uncalibrated Epipolar Constraint). *If the camera has an unknown calibration described by a non-singular matrix $A \in \mathbb{R}^{3\times 3}$, the epipolar constraint in the above theorem becomes:*

$$\boldsymbol{x}_2^T A^{-T}\widehat{T}RA^{-1}\boldsymbol{x}_1 = 0. \tag{30}$$

The matrix $F = A^{-T}\widehat{T}RA^{-1}$ is called fundamental matrix as in the Euclidean case.

In the Euclidean case, the epipolar constraint essential states the fact that the two optical centers and the 3-D point p being observed are coplanar. Another way to say this is that these three points form a triangle, which in turn determines a plane. Then in the more general cases, the epipolar constraint simply states the common fact that the two optical centers and the 3-D point p being observed must form a "geodesic triangle". This is illustrated in Figure 6.

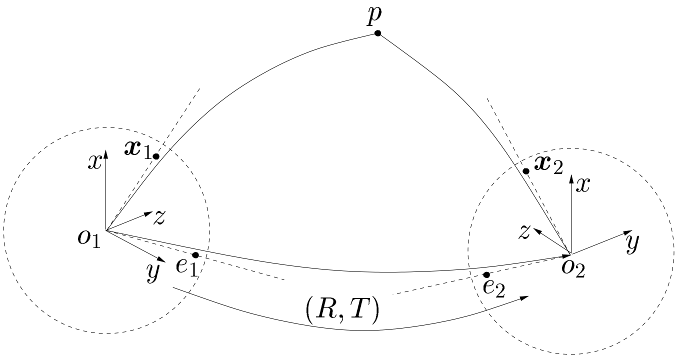

Figure 6. Geodesic triangle formed by two optical centers o_1, o_2 and a point p in the scene. $\boldsymbol{x}_1, \boldsymbol{x}_2$ are the two images of the point p. e_1, e_2 are the corresponding epipoles.

Comment 3 (Periodic Geodesics). *The condition* $\sin(\eta\rho) \neq 0$ *is equivalent to the condition that the translation* $T \neq 0$ *in the Euclidean case. The reason is when* $\sin(\eta\rho) = 0$*, we have* $\exp(X) = I_4$*, i.e. the motion is equivalent to the identity transformation on* M*. In spaces of constant curvature, we may have* $\sin(\eta\rho) = 0$ *without* $T = 0$*. This occurs only when the curvature* r *is positive, i.e. the space is spherical. If so, let* $\eta = 2k\pi\sqrt{r} \in \mathbb{R}, k = 1, 2, \ldots,$ *we then have* $\sin(\eta\rho) = \sin(2k\pi) = 0$*. This implies that a translation of distance* $2\pi\sqrt{r}$ *along the geodesics (big circles) in a spherical space of radius* $\sqrt{r}$ *is equivalent to the identity transformation – you travel back to the initial position by circling around the globe once. One can easily verify this on a 2 dimensional sphere* $\mathbb{S}^2$ *or a circle* $\mathbb{S}^1$.

As in the Euclidean case, using the epipolar constraint, the essential matrix $E = \widehat{T}R$ can be estimated (up to a scale) from more than eight image correspondences $\{(\boldsymbol{x}_1^j, \boldsymbol{x}_2^j)\}_{j=1}^n, n \geq 8$ in general positions using linear or nonlinear estimation schemes. The rotation matrix R and the translation vector T can further be recovered from decomposing the essential matrix E (see (Ma, Košecká and Sastry, 1999; Maybank, 1993) for the details). In the uncalibrated case, the fundamental matrix F is first recovered and the unknown calibration A can be solved from the well-known Kruppa's equations (Ma, Vidal, Košecká and Sastry, 1999) or other self-calibration methods.

3.3.2. *Triangulation Using Absolute Trigonometry*

Knowing the motion parameters (R, T), the next problem is how to reconstruct the scale information from images, which includes the scale λ of the point p with respect to its image $\boldsymbol{x}$, the distance of the translational motion along T and, if possible, the constant curvature $1/r$ of the space M. But we will soon see, not both the curvature and the scales can be uniquely determined from image measurements. We here first demonstrate the main ideas for two calibrated images.

To simplify the notation, in this section, we assume that the image $\boldsymbol{x}$ of a point p is always normalized, i.e. $\|\boldsymbol{x}\| = 1$ (in the Euclidean case, this corresponds to the so-called spherical projection). Suppose the distance from p to the optical center o is $\eta \in \mathbb{R}_+$. Then the homogeneous coordinate of p is given in terms of $\boldsymbol{x}$ and η by:

$$p = \begin{bmatrix} \rho^{-1}\sin(\eta\rho)\boldsymbol{x} \\ \cos(\eta\rho) \end{bmatrix} \quad \in \mathbb{R}^4. \tag{31}$$

Consequently, the scale λ of p with respect to $\boldsymbol{x}$ is given by $\lambda = \rho^{-1}\sin(\eta\rho)$. To differentiate from the scale λ, the distance quantity η will be called the *depth* of the point p with respect to the image $\boldsymbol{x}$.

Let η_1 and η_2 be the depths of the point p with respect its two images $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$ taken by the camera at two positions, respectively. Suppose the camera motion $g \in G$ is specified by the rotation $R \in SO(3)$, the translation direction $T \in \mathbb{S}^2$ and the distance of translation γ (as in the preceding section). Then the first equation in (26) yields:

$$\begin{aligned} \rho^{-1}\sin(\eta_2\rho)\boldsymbol{x}_2 \;&=\; \left[I_3 + (\cos(\gamma\rho) - 1)TT^T\right] R\rho^{-1}\sin(\eta_1\rho)\boldsymbol{x}_1 \\ &\quad + \cos(\eta_1\rho)\rho^{-1}\sin(\gamma\rho)T. \end{aligned} \tag{32}$$

This is the *coordinate transformation formula in spaces of constant curvature*.

Comment 4. *Although seemingly a little complicated, the above equation is no more than a natural generalization of the Euclidean coordinate transformation formula which people are familiar with. Notice that when the curvature $1/r$ goes to zero, so does $\rho = \sqrt{1/r}$. Since*

$$\lim_{\rho\to 0}\cos(\eta_i\rho) = 1, \quad \lambda_i = \lim_{\rho\to 0}\rho^{-1}\sin(\eta_i\rho) = \eta_i, \tag{33}$$

in the Euclidean case (32) simply becomes:

$$\lambda_2\boldsymbol{x}_2 = R\lambda_1\boldsymbol{x}_1 + \gamma T. \tag{34}$$

That it, in the limit case, the scale λ and the depth η are the same; and the equation (32) gives rise to the familiar Euclidean coordinate transformation formula. Naturally, to reconstruct structure in spaces of constant curvature, the equation (32) has to be exploited instead.

Notice that equation (32) is homogeneous in the scale of ρ. Since the quantities η_1, η_2 and γ are all multiplied with ρ, they can only be determined with respect to an arbitrary scale of ρ. Thus in the case of spaces of constant curvature, we may normalize everything with respect to the scale of the curvature: if $r > 0$, let $\rho = 1$; if $r < 0$, let $\rho = i = \sqrt{-1}$. That is, now the space M has constant sectional curvature of either $+1$ or -1. Then (32) respectively becomes:

$$\begin{aligned}
\sin(\eta_2)\boldsymbol{x}_2 &= \left[I_3 + (\cos(\gamma) - 1)TT^T\right] R \cdot \sin(\eta_1)\boldsymbol{x}_1 \\
&\quad + \cos(\eta_1)\sin(\gamma)T, \quad \rho = 1; \\
\sinh(\eta_2)\boldsymbol{x}_2 &= \left[I_3 + (\cosh(\gamma) - 1)TT^T\right] R \cdot \sinh(\eta_1)\boldsymbol{x}_1 \\
&\quad + \cosh(\eta_1)\sinh(\gamma)T, \quad \rho = i.
\end{aligned}$$

These two equations correspond to coordinate transformations in (normalized) spherical and hyperbolic spaces, respectively.

Proposition 4. *In a space of constant curvature, the curvature of the space is not recoverable from multiple images unless the distance between points are known a priori.*

From the preceding section, we know R and T can be estimated from epipolar constraints. The problem left is to reconstruct η_1, η_2 and γ. Suppose that the two optical centers of the camera are o_1 and o_2. A *geodesic triangle* is formed by the three points (o_1, o_2, p), see Figure 7. The angle A is given by

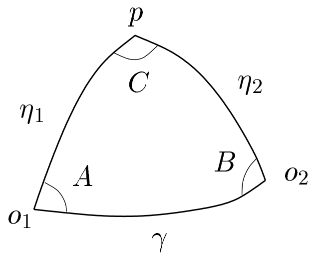

Figure 7. Geodesic triangle formed by two optical centers o_1, o_2 and a point p in the scene. For this triangle, we have $a = \eta_2, b = \eta_1$ and $c = \gamma$.

the angle between the two vectors $\boldsymbol{x}_1$ and $-R^T T$; B is given by the angle between $\boldsymbol{x}_2$ and T. Unlike the Euclidean case, we here cannot directly compute the angle C from A, B since the sum of $A + B + C$ is not necessarily π in the case of a Riemannian manifold. According to the Gauss-Bonnet Theorem:

$$A + B + C - \pi = K \cdot \text{area}(\triangle ABC) \tag{35}$$

since here the space has a constant Gauss curvature $K = \pm 1$ after normalization.

The problem of determining lengths η_1, η_2, γ from such a triangle is usually referred to as *triangulation* as in the Euclidean case (Hartley and Sturm, 1997). In the Euclidean case, one may directly use above coordinate transformation formula (34) to formulate linear least square type of objective functions for estimating depths η and γ (Ma, Košecká and Sastry, 1999). However, in the non-Euclidean case, such objective functions are much more complicated in the unknown η and γ for obvious reasons.

In stead of directly solving for all the unknown variables simultaneously, let us first try to identify the minimum number of constraints on such a triangle based on the well-known trigonometry in spaces of constant curvature. They are the so-called *Bolyai's law of sine* and *law of cosine* (in the case of absolute geometry), which are summarized in (Hsiang, 1995). Define functions:

$$\alpha(x) = \begin{cases} \sin(x), & \rho = 1, \\ \sinh(x), & \rho = i, \end{cases} \qquad \beta(x) = \begin{cases} \cos(x), & \rho = 1, \\ \cosh(x), & \rho = i. \end{cases}$$

The next proposition follows from (Hsiang, 1995) as a special case:

Proposition 5 (Laws of Absolute Trigonometry). *Consider a geodesic triangle $\triangle ABC$ in a space M of constant curvature ± 1, and let a, b, c be the lengths of the opposite sides of angles A, B, C respectively. Then we have:*

1. *Bolyai's law of sine:*

$$\frac{\sin(A)}{\alpha(a)} = \frac{\sin(B)}{\alpha(b)} = \frac{\sin(C)}{\alpha(c)}, \tag{36}$$

2. *Bolyai's law of cosine:*

$$\begin{aligned} \alpha(a)\alpha(b)\cos(C) &= \beta(c) - \beta(a)\beta(b), \\ \alpha(b)\alpha(c)\cos(A) &= \beta(a) - \beta(b)\beta(c), \\ \alpha(c)\alpha(a)\cos(B) &= \beta(b) - \beta(c)\beta(a). \end{aligned} \tag{37}$$

In our case only the quantities $\sin(A), \sin(B), \cos(A), \cos(B)$ in the above equations (36) and (37) are known to us. Either the sine law or the cosine law gives only two independent algebraic constraints on three unknowns $a = \eta_2, b = \eta_1, c = \gamma$. Hence it is in general impossible to uniquely determine η_1, η_2, γ.[13] Like the Euclidean case, this corresponds to the well-known fact that the structure can only be reconstructed up to a universal scale. Hence without loss of generality, we can further normalize the distance γ of translation such that

$$\alpha(\gamma) = 1, \quad \text{or equivalently} \quad \beta(\gamma) = 0. \tag{38}$$

The cosine law gives $\cos(C) = -\cos(A)\cos(B)$ and so we know $\sin(C)$ too. Consequently the depths η_1, η_2 are given by:

$$\alpha(\eta_1) = \frac{\sin(A)}{\sin(C)}, \quad \alpha(\eta_2) = \frac{\sin(B)}{\sin(C)} \tag{39}$$

from the sine law since $\alpha(\gamma) = 1$.

3.4. Geometry and Reconstruction from Multiple Views

It is already known that in the Euclidean case, m images of a point satisfy certain *multilinear* constraints besides the bilinear epipolar constraints. Similar constraints exist in the case of spaces of constant curvature. We here give a complete study of those constraints through the use of the so-called *rank condition*. Relationships among different types of constraints can also be easily revealed this way.

Suppose $\boldsymbol{x}_i \in \mathbb{R}^3, i = 1, 2, \ldots, m$ are m images of the same point p with respect to the camera at m different positions (or vantage points). Suppose the relative motion between the i^{th} and 1^{th} positions is $g_i \in G, i = 1, 2, \ldots, m$. Let $\lambda_i \in \mathbb{R}_+, i = 1, 2, \ldots, m$ be the scales of p w.r.t. the images $\boldsymbol{x}_i, i = 1, 2, \ldots, m$. Then we have the following equation:

$$\begin{bmatrix} \boldsymbol{x}_1 & 0 & \cdots & 0 \\ 0 & \boldsymbol{x}_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \boldsymbol{x}_m \end{bmatrix} \begin{bmatrix} \lambda_1 \\ \lambda_2 \\ \vdots \\ \lambda_m \end{bmatrix} = \begin{bmatrix} APg_1 \\ APg_2 \\ \vdots \\ APg_m \end{bmatrix} p. \tag{40}$$

Let us call the matrix $\Pi_i \doteq APg_i \in \mathbb{R}^{3\times 4}$ as the i^{th} *projection matrix*. Without loss of generality, we may always assume that the first projection matrix $\Pi_1 = APg_1$ is of the standard form $[I, 0] \in \mathbb{R}^{3\times 4}$.[14] In general the i^{th} projection matrix is of the form:

$$\Pi_i = [AW_iR_i, Ay_i] \tag{41}$$

where W, y are given in (27). For simplicity, we call the first three columns of Π_i as $U_i \doteq AW_iR_i \in \mathbb{R}^{3\times 3}$ and the last column as $V_i \doteq Ay_i \in \mathbb{R}^3$. Now define a so-called *multiple view matrix* M^a to be

$$M^a = \begin{bmatrix} \widehat{\boldsymbol{x}_2}U_2\boldsymbol{x}_1 & \widehat{\boldsymbol{x}_2}V_2 \\ \widehat{\boldsymbol{x}_3}U_3\boldsymbol{x}_1 & \widehat{\boldsymbol{x}_3}V_3 \\ \vdots & \vdots \\ \widehat{\boldsymbol{x}_m}U_m\boldsymbol{x}_1 & \widehat{\boldsymbol{x}_m}V_m \end{bmatrix} \in \mathbb{R}^{3(m-1)\times 2}. \tag{42}$$

The superscript a in M^a indicates *absolute geometry*. Similar to the Euclidean case (Ma, et. al., 2001), in spaces of constant curvature, we also have:

Theorem 2 (Multiview Rank Condition). *Consider m images $\{\boldsymbol{x}_i\}_{i=1}^m \in \mathbb{R}^3$ of a point p in a space M of constant curvature. The multiple view matrix M^a defined above satisfies*

$$\boxed{0 \leq rank(M^a) \leq 1.} \tag{43}$$

The proof is essentially the same as in the Euclidean case (Ma, et. al., 2001). For the same reasons as in the Euclidean case, it is straightforward to see that non-trivial constraints given by the above rank condition are either *bilinear* or *trilinear* in the image coordinates $\boldsymbol{x}_i$'s:[15]

$$\boldsymbol{x}_i^T \widehat{V}_i U_i \boldsymbol{x}_1 = 0, \quad \widehat{\boldsymbol{x}_i}(V_i \boldsymbol{x}_1^T U_j^T - U_i \boldsymbol{x}_1 V_j^T)\widehat{\boldsymbol{x}_j} = 0. \tag{44}$$

As we see, the bilinear type gives exactly the epipolar constraint we derived before (30).

Corollary 2 (Linear Relationships among Multiple Views of a Point). *For any given m images corresponding to a point $p \in M$ relative to m camera frames, that the matrix M^a is of rank no more than 1 yields the following:*

1. *Any algebraic constraint among the m images can be reduced to only those involving 2 and 3 images at a time. Formulae of these bilinear and trilinear constraints are given in (44) respectively. There is no other relationship among point features in more than three views.*

2. *For given m images of a point, all the triple-wise trilinear constraints algebraically imply all pairwise bilinear constraints, except in the degenerate case in which the pre-image p lies on the geodesic through optical centers o_1, o_i for some i.*

It is also easy to see that the kernel of the multiple view matrix M^a is

$$M^a \begin{bmatrix} \lambda_1 \\ 1 \end{bmatrix} = 0 \tag{45}$$

where λ_1 is exactly the scale of the point p relative to the first camera frame. Hence the multiple view matrix associated to m images exactly captures the scale (or depth) information (of a point) that is missing in a single image but encoded in multiple ones. The kernel of M^a is unique except when rank$(M^a) = 0$, i.e. $M^a = 0$. The latter corresponds to a rare configuration where all camera centers $o_1, \ldots, o_m$ and the point p lie on the same geodesic. We hence have the following statement regarding a geometric interpretation of the multiple view matrix:

Corollary 3 (Uniqueness of the Pre-image). *Given m vectors on the image planes with respect to m camera frames, they correspond to the same point in the 3-D space M if the rank of the M^a matrix is 1. If its rank is 0, the point is determined up to the geodesic where all the camera centers must lie on.*

This is illustrated in Figure 8. The reader may have been aware that above results are very much consistent with what we have known in multiple view geometry for the Euclidean space. The rank conditions on the multiple view

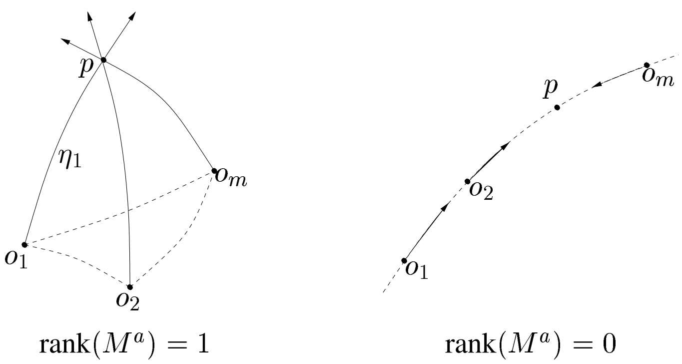

Figure 8. Two cases corresponding to the two rank values of M^a. Left: a generic case; Right: a degenerate case.

matrix also applies to line or plane features in the Euclidean case. We should therefore expect a similar story holds in the general case. We here omit the detail for simplicity.

From the above study, we see that the distinction of multiple view geometry between Euclidean and non-Euclidean spaces is very subtle. They all obey the same projection model:

$$\lambda_i \boldsymbol{x}_i = \Pi_i p \tag{46}$$

except that the internal structure of the projection matrix $\Pi_i \in \mathbb{R}^{3\times 4}$ may be different. This has therefore revealed a very interesting and important fact: The same set of parameters for the projection matrices

$$\Pi \doteq \begin{bmatrix} \Pi_1 \\ \Pi_2 \\ \vdots \\ \Pi_m \end{bmatrix} = \begin{bmatrix} AW_1R_1 & Ay_1 \\ AW_2R_2 & Ay_2 \\ \vdots & \vdots \\ AW_mR_m & Ay_m \end{bmatrix} \in \mathbb{R}^{3m\times 4} \tag{47}$$

can have different interpretations. They are all geometrically meaningful. For example, we know the above Π can be interpreted as the projection matrices of an uncalibrated camera with constant calibration A moving in a space of constant curvature. It can also be interpreted as an uncalibrated camera with *time-varying* calibration AW_i moving in a Euclidean space. Then R_i and $W_i^{-1}y_i$ become the rotation and translation of the camera motion. Hence essentially

Corollary 4 (Equivalence of Imaging Systems). *Taking images of an object in a non-Euclidean space is equivalent to taking images of (the same object) in a Euclidean space (probably at a different set of vantage points) and introducing on each image an unknown linear transformation that depends on the vantage point.*

According to this, reconstruction of a 3 dimensional scene from multiple images in general is ambiguous. It only becomes well-conditioned if we have sufficient a priori knowledge on the space, the camera calibration, or the camera motion.

This extra ambiguity and complexity in recovering the projection matrix $\Pi_i = [AW_iR_i, Ay_i]$ make the problem of obtaining a global reconstruction from multiple images of multiple points much harder in the non-Euclidean case. For instance, conventional projective factorization and stratification methods (Hartley and Zisserman, 2000) designed for the Euclidean case no longer apply to the non-Euclidean case even in the simplest case $A = I$. However, factorization methods based on the above rank condition (43) work just the same, see (Ma, et. al., 2001). We here omit the detail for simplicity.

4. Extensions and Applications to Spaces of Non-constant Curvature

We all know light travels a path of the shortest distance. However, in general, "a path of the shortest distance" does not necessarily mean a straight line. For instance, it is well-known in physics that when light or seismic wave travels across inhomogeneous media, its trajectory is bent according to the Snell's law of sine, as shown in Figure 9:

$$\boxed{n\sin(\theta) = \text{constant}} \tag{48}$$

where θ is the incidence angle and n is the index of refraction. n is typically the ratio c/v between the speed of light in vacuum and that in the media. Figure 9 demonstrates two cases when n is piecewise constant or varies smoothly. The latter case usually occurs when light travels through a depth of sea water with different mineral density or seismic waves through the earth. Geometrically we now know that the reason for such refraction can be explained as a different distribution of material introduces a different (Riemannian) metric to the space. Such metric in general induces a non-zero (sectional) curvature to the space and makes it no longer Euclidean. Nonetheless, the light always travels the shortest distance with respect to the new metric – the Snell's law simply describes what these shortest distance paths are in some special cases. Hence the geometric framework developed in this paper can certainly be used to unify the study of vision and imaging problems associated to these phenomena. For instance, restriction of the multiple view geometry introduced in this paper to a 2 dimensional spherical space can be used to locate the center of earthquake (on the surface of earth) by observing seismic waves from multiple observation posts far away.

This paper only studies a rather idealistic case when the sectional curvature of the space is constant in all directions. Such an idealization undoubtly

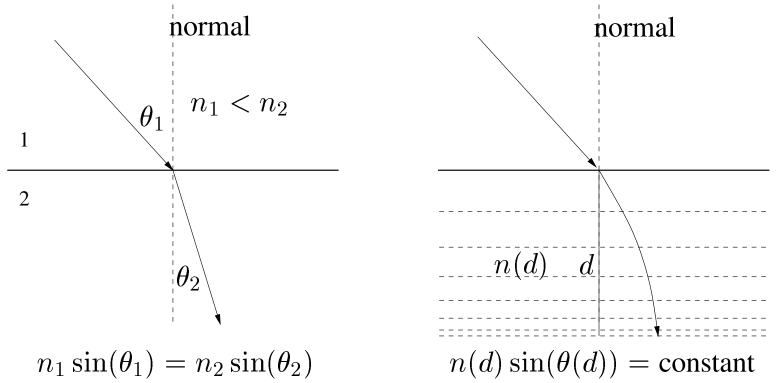

Figure 9. Snell's law of sine. Left: A ray of light travels through two different media; Right: A ray of light travels through a media with an index of refraction n as an increasing continuous function of the depth d.

will limit its application to real-world situations. However, it is an important conceptual step towards any further development of more realistic imaging and vision models for non-Euclidean spaces which are of much more practical importance. For example, in the case of gravitational lensing (see Figure 2), the earth-galaxy-quasar geometry can be approximated as a space of piece-wise constant curvature as shown in Figure 10.

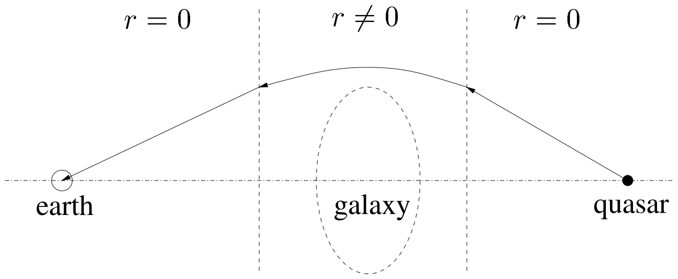

Figure 10. An approximate space model for the earth-galaxy-quasar geometry. Around the large galaxy, the space has a non-zero curvature due to the gravitation of a large mass.

As one may have recognized that essentially the camera model for a space of non-zero curvature is to allow a *non-Euclidean motion* (described by the group G) from one camera frame to another. A similar scenario arises when multiple images of a dynamical scene are considered. Although it is shown in (Kun, Fossum and Ma, 2001) that a dynamical scene can usually be embedded in a higher dimensional Euclidean space, the transformation between projection matrices is typically *non-Euclidean*, which is imposed not by a curvature but by the nature of dynamics in the scene. Therefore, the proposed study of the more generalized model (4) other than the conventional ones (3) becomes necessary again when one wants to generalize classic multiple view geometry to dynamical scenes.

5. Conclusions

In this paper, we have generalized basic results in multiple view geometry for Euclidean space to spaces of constant curvature. A uniform treatment is possible because a unified homogeneous representation of these spaces exists and the isometry groups of these spaces are naturally embedded in $GL(4, \mathbb{R})$. Consequently, multiple view geometry for spaces of constant curvature is remarkably similar to that for Euclidean space. In particular, epipolar constraint remains exactly the same and so do conditions for dependency among multilinear constraints. This allows us to extend most motion recovery algorithms previously developed for Euclidean space to spherical and hyperbolic spaces with little change. As for the triangulation problem, the three dimensional structure can only be reconstructed up to a universal scale, the same as the Euclidean case. Moreover, without knowing the scale, the curvature of the space cannot be recovered from image measurements at all – only its sign can be detected.

When the sectional curvature of a Riemannian manifold in all directions is approximately the same, it can be locally modeled as one of the space forms studied in this paper. The multiple view geometry that we have developed may provide a good approximation to the vision problem in such a manifold. However, it remains a question whether results such as epipolar constraint still generalize to more general classes of Riemannian manifolds (for example, symmetric spaces of rank 1); and how the triangulation method needs to be modified. At this point, little is known about vision beyond spaces of constant curvature. It remains an open problem for future investigation.

As we have illustrated before, a general theory of multiple view geometry in general Riemannian manifolds can be useful for seismic or astronomical purposes, where typically the curvature of a space can no longer be neglected due to either a large scale or a lack of homogeneousness in spatial property. However, there has not yet been much effort to study geometric properties of such spaces (or manifolds) from a vision perspective much less we know about vision in a space-time where *relativity* plays a significant role, i.e. when the speed of light can no longer be assumed to be infinite. In any case, a large part of relationships among vision, motion, and space is yet to be investigated in a more general geometric setting.

Appendix

A. Euclidean Space as a Space of Constant Curvature

Proposition 1 requires the curvature parameter $r \neq 0$ hence only the spherical and hyperbolic spaces were considered. However, the Euclidean case can be regarded as the limit case when r goes to infinite, i.e. the sectional curvature $1/r$ goes to zero.

When $r = \infty$, a point in $\mathbb{R}^4$ which satisfies the quadratic form (5) always has the form $[x_1, x_2, x_3, 1]^T \in \mathbb{R}^4$. This is just the homogeneous representation of the 3 dimensional Euclidean space $\mathbb{E}^3$, see (Murray, Li and Sastry, 1994). Then the condition (7) has become:

$$w^2 = 1, \quad z = 0, \quad W^T W = I_3, \quad y \in \mathbb{R}^3. \tag{49}$$

Thus the group G is just the Euclidean group $E(3)$. In particular, the special Euclidean group $SE(3)$ with elements:

$$g = \begin{bmatrix} R & T \\ 0 & 1 \end{bmatrix} \quad \in \mathbb{R}^{4\times 4}$$

with $R \in SO(3)$ and $T \in \mathbb{R}^3$ is an orientation-preserving subgroup of the Euclidean group $G = E(3)$. As we already know, $SE(3)$ represents the *rigid body motion* in $M = \mathbb{E}^3$.

When $r = \infty$, the Lie algebra $se(3)$ of $SE(3)$ or $e(3)$ of $E(3)$ then has the form given in (9) with the condition $c = 0$. In robotics literature (Murray, Li and Sastry, 1994), an element of this Lie algebra is usually represented as *twist*:

$$\xi = \begin{bmatrix} \widehat{\omega} & v \\ 0 & 0 \end{bmatrix} \quad \in \mathbb{R}^{4\times 4}$$

where $\omega, v \in \mathbb{R}^3$ and $\widehat{\omega}$ is the skew-symmetric matrix associated with $\omega = [\omega_1, \omega_2, \omega_3]^T$ such that $\widehat{\omega} v = \omega \times v, \forall v \in \mathbb{R}^3$:

$$\widehat{\omega} = \begin{bmatrix} 0 & -\omega_3 & \omega_2 \\ \omega_3 & 0 & -\omega_1 \\ -\omega_2 & \omega_1 & 0 \end{bmatrix} \quad \in \mathbb{R}^{3\times 3}. \tag{50}$$

According to Proposition 2, a geodesic in $\mathbb{E}^3$ is given by the form:

$$\exp\left(t \begin{bmatrix} 0 & v \\ 0 & 0 \end{bmatrix}\right) = \begin{bmatrix} I_3 & vt \\ 0 & 0 \end{bmatrix} \quad \in \mathbb{R}^{4\times 4} \tag{51}$$

which is exactly a translation in the direction of v.

From the above discussion, the Euclidean space can be treated as a limit case of general spaces of constant curvature characterized by Proposition 1. Because of this, the vision theory for Euclidean space should also be a limit case of vision theory for general spaces of constant curvature.

Acknowledgements

The author likes to thank Professor Alan Weinstein of the Mathematics department at UC Berkeley for his insightful comments on geometric properties of symmetric spaces.

Notes

[1] $GL(n, \mathbb{R})$ is the group of all non-singular $n \times n$ real matrices; $A(n-1, \mathbb{R})$ is the group of $n \times n$ matrices of the form $\begin{bmatrix} M_1 & M_2 \\ 0 & 1 \end{bmatrix}$ with $M_1 \in GL(n-1, \mathbb{R})$ and $M_2 \in \mathbb{R}^{n-1}$.

[2] Nasa news: http://science.nasa.gov/newhome/headlines/ast14may99_1.htm

[3] Source: http://antwrp.gsfc.nasa.gov/apod/ap980111.html

[4] Source: http://science.nasa.gov/newhome/headlines/ast14may99_1.htm

[5] A subgroup of G which fixes a point of M.

[6] A *frame bundle* is the set of all coordinate frames associated to each point of the manifold M.

[7] For a general symmetric space, G_H is not necessarily the full orthogonal group but it is for spaces of constant curvature.

[8] In fact, the orthonormal frame bundle of M is isomorphic to G as a principle G_H bundle.

[9] A Lie algebra of a Lie group is the tangent space of the group at its identity element.

[10] A Lie bracket of two matrices $A, B \in \mathbb{R}^{n \times n}$ is defined as $[A, B] = AB - BA \in \mathbb{R}^{n \times n}$.

[11] An exponential of a matrix $X \in \mathbb{R}^{n \times n}$ is defined as

$$\exp(X) = I + X + \frac{X^2}{2!} + \cdots + \frac{X^n}{n!} + \cdots .$$

When X belongs to a Lie algebra of a matrix Lie group G, then this exponential map coincides with the one defined for G as a Riemannian manifold with its canonical metric if there exists one.

[12] In the hyperbolic case, G_T acts at least transitively on each one of the two disconnected hyperbolic surfaces. That is enough for our purpose here.

[13] In fact, there is a family of infinitely many solutions.

[14] By doing so, we essentially choose the first camera frame to be the reference.

[15] In the Euclidean case, these constraints are also sometimes referred as bifocal and trifocal tensorial constraints in the Computer Vision literature.

References

Faugeras, O. Stratification of three-dimensional vision: projective, affine, and metric representations. *Journal of the Optical Society of America*, 12(3):465–84, 1995.

Huang, K. Fossum, R., and Ma, Y. Generalized rank conditions in multiple view geometry with applications to dynamical scenes. submitted to *European Conference on Computer Vision*, 2002.

Hartley, R. and Sturm, P. Triangulation. *Computer Vision and Image Understanding*, 68(2):146–57, 1997.

Hartley, R. and Zisserman, A. *Multiple View Geometry in Computer Vision*. Cambridge University Press, 2000.

Heyden, A. and Åström, K. Algebraic properties of multilinear constraints. *Mathematical Methods in Applied Sciences*, 20(13):1135–62, 1997.

Hsiang, W.-Y. Absolute geometry revisited, Center for Pure and Applied Mathematics, University of California at Berkeley. *PAM-628*, 1995.

Kobayashi, S. and Nomizu, T. *Foundations of Differential Geometry: Volume I and Volume II*. John Wiley & Sons, Inc., 1996.

Ma, Y., Košecká, J. and Sastry, S. Optimization criteria and geometric algorithms for motion and structure estimation. *accepted by International Journal of Computer Vision*, 2001.

Ma, Y., Huang, K., Vidal, René, Košecká, J. and Sastry, S. Rank conditions on the multiple view matrix. submitted to *International Journal of Computer Vision* , 2001.

Ma, Y., Vidal, R., Košecká, J. and Sastry, S. Kruppa's equation revisited: its degeneracy and renormalization. In *Proceedings of ECCV*, Dublin, Ireland, 2000.

Maybank, S. *Theory of Reconstruction from Image Motion*. Springer-Verlag, 1993.

Murray, R., Li, Z. and Sastry, S. *A Mathematical Introduction to Robotic Manipulation*. CRC press Inc., 1994.

Schneider, P., Ehlers, J., and Falco, E.E. (ed.), *Gravitational Lenses*. Springer-Verlag, Berlin, 1992.

Weinstein, A. Mathematics Department, UC Berkeley. *Personal communications*, 2000.

Wolf, J. A. *Spaces of Constant Curvature*. Publish or Perish, Inc., 5th edition, 1984.

List of Figures