A LIE THEORETIC APPROACH TO STRUCTURE AND MOTION IN COMPUTER VISION [1]

Yi Ma * **Omid Shakernia** * **Jana Košecká** *
Shankar Sastry *

* {*mayi,omids,janka,sastry*}*@robotics.eecs.berkeley.edu*
Electronics Research Laboratory
University of California at Berkeley
Berkeley, CA 94720-1174, USA

Abstract: There has been an increasing interest in applying computer vision in robot control such as vision guided navigation, manipulation, and object recognition. A combination of a robot's rigid body motion and a camera's perspective projection has brought us a new geometric subject to study, the so called *multiview geometry* in computer vision literature. In this paper, we propose a new approach to this subject based on mathematical tools commonly used in robotics such as linear algebra, differential geometry and Lie group theory. Based on the geometric model of a vision system, main results of multiview geometry are outlined, including reconstruction theory of camera motion, scene structure and camera self-calibration. Potential applications of multiview geometry in robot control are also presented.

Keywords: multiview geometry, structure from motion, camera self-calibration, visual servoing.

1. INTRODUCTION

There has been an increasing interest in the use of computer vision systems in the navigation of autonomous mobile agents, such as an unmanned vehicle or a helicopter (Koo *et al.*, 1998; Kosecka *et al.*, 1997). This interest has motivated us to study computer vision from a control-theoretic point of view — the character of an "idealized" camera in a "low level" feedback loop.

The first step is to understand the geometry of an idealized vision system (without introducing kinematics or dynamics yet). In computer vision literature, this subject is called multiview geometry. At a certain level, multiview geometry is governed by two fundamental transformations: the perspective projection which models the input-output characteristic of a camera; and the transformation group (a Lie group) which models the motion of the camera. [2] Due to prime interest of the perspective projection, in computer vision literature, projective geometry has been developed as one of the main mathematical tools in the study of multiview geometry. Besides its own mathematical importance, projective geometry provides an abstract framework in which certain properties of perspective images can be naturally studied (Faugeras, 1993).

In the projective geometry framework, a problem is usually reduced to one of solving a set of homogeneous algebraic equations; classical algebraic

[1] This work is supported by ARO under the MURI grant DAAH04-96-1-0341.

[2] For simplicity, we here assume the rigidity of both the camera and the objects. When the rigidity assumption is violated, to some extent, it may be treated as a vision problem in a higher dimensional (Riemannian) space.

geometry methods are then applied to analyze and solve these equations. This approach is very powerful in the sense of finding a solution, but one of its main drawbacks is that intuitive geometric nature of the problem can be easily lost in such a formulation. For instance, it is known that, in a three dimensional Euclidean space, whether a camera is calibrated or not, the transformation group describing the camera motion is isomorphic to the special Euclidean group $SE(3)$. In the projective geometry framework, this group is "artificially" extended to the general linear group $GL(4)$, hence less geometric invariants can be (directly) exploited to solve the structure and motion reconstruction problem. Moreover, to study vision in a dynamical setting, the effect of camera dynamics on image features needs to be described precisely. Then the Euclidean group and perspective projection become equally important, and a mathematical framework other than projective geometry is therefore needed to combine them together.

In this paper, we propose a Lie theoretic framework based on notions from differential geometry (although techniques needed to solve concrete problems usually only involve linear algebra). We will show how, in this framework, main problems in multiview geometry, *i.e.* recovery of motion, structure and calibration, can be solved in a rather unified fashion, and, due to the generality of our formulation, the obtained results also apply to non-Euclidean spaces.

The next step is to study vision in a dynamical setting and this can be regarded as a "natural extension" of the multiview geometry — the combination of the group of camera motion with the perspective projection is now replaced by the combination of camera kinematics (or dynamics) with the perspective projection. One possible way to make this combination is by *lifting* camera kinematics (or dynamics) onto the image plane, see (Ma *et al.*, to appear). Potential applications of multiview geometry in robot control will be discussed in more detail in the final section of this paper.

2. MATHEMATICAL MODEL OF A GEOMETRIC VISION SYSTEM

In this section, we give an axiomatic formulation of the mathematical model of a geometric vision system (in a Riemannian manifold). Although this model seems to be given in a rather abstract manner, it is a natural generalization of the conventional camera model in a Euclidean space, and such a generalization allows us to fully realize the geometric nature of a computer vision system, in a very concise and precise way.

Let us consider a (connected) Riemannian manifold (M, g), *i.e.* a differentiable manifold equipped with a positive definite 2-form g as its metric. If the reader is not familiar with differential geometry, he or she may simply view (M, g) as the Euclidean space $\mathbb{R}^3$ with its standard inner product metric. In this paper, we will be mostly interested in three 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. (Space). M is a complete and orientable Riemannian manifold. G is the orientation-preserving subgroup of the isometry group of M. This group 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 geodesics with constant speed. For simplicity, we may assume this speed to be infinite.

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

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

Note that, in the Euclidean case, the first four assumptions are equivalent to the conventional *perspective* or *spherical* projection model. Since our approach is essentially coordinate free, these two models can be treated as the same. In general, the above model can be illustrated in the Figure 1.

The Lie group G which models the motion of the camera is obtained in the model as being the isometry group of M. In fact the relation between G and M is symmetric: letting H be the isotropy subgroup[3] of G, then the manifold M is simply the quotient space G/H. The Riemannian metric g on M can be derived from the canonical metrics of G and H by this quotient. In practice, this viewpoint is far more useful than the above axiomatic definition since, as we will soon see, interesting manifolds are usually given

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

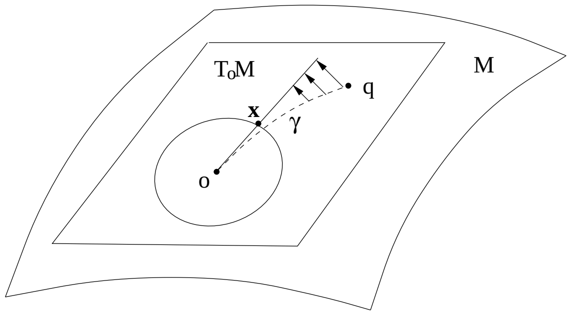

Fig. 1. The curve γ is the geodesic connecting o and q; arrows mean the inverse of the exponential map $\exp : T_oM \to M$; $\mathbf{x}$ then represents the image of the point q with respect to a camera centered at the point o.

as submanifolds of an Euclidean space which are invariant under the action of certain Lie groups G. Therefore, geometric properties of a vision system in such manifolds are uniquely determined by the structure of G.

For example, consider that a Lie group G is given as the subgroup of $GL(4)$ (general linear transformation group of $\mathbb{R}^4$) which preserves the quadratic form:

$$x_1^2 + x_2^2 + x_3^2 + rx_4^2$$

where $r \in \mathbb{R}$ is a nonzero constant. Then G can be represented as:

$$G = \{g \in \mathbb{R}^{4\times 4} \,|\, g^T Q g = Q\} \tag{1}$$

where Q is the 4×4 symmetric matrix:

$$Q = \begin{pmatrix} I_3 & 0 \\ 0 & r \end{pmatrix}.$$

Then G preserves a three dimensional hypersurface M in $\mathbb{R}^4$ given by:

$$x_1^2 + x_2^2 + x_3^2 + rx_4^2 = r.$$

If we impose a Riemannian metric g on M by restricting the following 2-form to M:

$$dx_1^2 + dx_2^2 + dx_3^2 + r\, dx_4^2,$$

then the manifold M is a three dimensional space of constant sectional curvature $1/r$ (Kobayashi and Nomizu, 1996). Note that when $r > 0$, M is a three dimensional sphere; when $r < 0$, M is a three dimensional hyperbolic space; when $r = \infty$, M is simply the three dimensional Euclidean space with all points represented in homogeneous coordinates $(x_1, x_2, x_3, 1)^T \in \mathbb{R}^4$. In the last case, the group G is simply $SE(3)$,[4] the special Euclidean group, represented however in homogeneous form (Murray *et al.*, 1993):

$$G = \left\{ \begin{pmatrix} R & p \\ 0 & 1 \end{pmatrix} \in \mathbb{R}^{4\times 4} \middle| R \in SO(3), p \in \mathbb{R}^3 \right\}$$

[4] Strictly speaking, G is the full Euclidean group $E(3)$ as defined in (1), but we always only pick the orientation-preserving component.

Then in spaces of constant curvature, the geometric camera model given above can be written uniformly as:

$$\lambda \mathbf{x} = APgq \tag{2}$$

where $\mathbf{x} \in T_oM \simeq \mathbb{R}^3$ is the image of a point $q \in M \subset \mathbb{R}^4$, $\lambda \in \mathbb{R}$ is a scalar giving the unknown distance of the point q away from the camera center o, $g \in G$ is the relative motion of the camera, P is the 3×4 matrix $(I_3, 0) \in \mathbb{R}^{3\times 4}$ which serves here as the inverse of the exponential map $\exp : T_oM \to M$, and $A \in SL(3)$ is a 3×3 matrix of determinant 1 which models the unknown camera calibration (in the case of calibrated camera, A is simply the identity I_3). For more detail about this equation, please refer to (Ma and Sastry, 1998).

One of the main topics of multiview geometry is *how to recover the unknown camera motion g, scene structure λ and camera calibration A from the image $\mathbf{x}$ only.* In the rest of this paper, we outline how this problem is solved in the Euclidean case although all the results in fact hold for spaces of constant curvature in general (Ma and Sastry, 1998).

3. MULTIVIEW GEOMETRY AND MOTION AND STRUCTURE RECOVERY

Suppose the camera takes images of a point q at m different views, we then have:

$$\lambda_i \mathbf{x}_i = A_i P g_i q, \quad i = 1, \ldots, m$$

where the camera calibration A_i's may vary from one view to another. Write these m equations in a matrix form:

$$\begin{pmatrix} \mathbf{x}_1 & 0 & \cdots & 0 \\ 0 & \mathbf{x}_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \mathbf{x}_m \end{pmatrix} \begin{pmatrix} \lambda_1 \\ \lambda_2 \\ \vdots \\ \lambda_m \end{pmatrix} = \begin{pmatrix} A_1 P g_1 \\ A_2 P g_2 \\ \vdots \\ A_m P g_m \end{pmatrix} q.$$

Now define the *image matrix* $\mathbf{X} \in \mathbb{R}^{3m\times m}$ and the *motion matrix* $\mathbf{A} \in \mathbb{R}^{3m\times 4}$ to be:

$$\mathbf{X} = \begin{pmatrix} \mathbf{x}_1 & 0 & \cdots & 0 \\ 0 & \mathbf{x}_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \mathbf{x}_m \end{pmatrix}, \quad \mathbf{A} = \begin{pmatrix} A_1 P g_1 \\ A_2 P g_2 \\ \vdots \\ A_m P g_m \end{pmatrix}.$$

The m columns of $\mathbf{X}$ are denoted by $\tilde{\mathbf{x}}_i, i = 1, \ldots, m$ and the four columns of $\mathbf{A}$ are denoted by a_1, a_2, a_3, a_4, respectively.

Theorem 1. (Multilinear Constraints). Consider m images $\{\mathbf{x}_i\}_{i=1}^m \in \mathbb{R}^3$ of a point $q \in M$, and the associated image matrix $\mathbf{X}$ and motion matrix $\mathbf{A} \in \mathbb{R}^{3m\times 4}$ as defined above. Then the column vectors of $\mathbf{X}$ and $\mathbf{A}$ satisfy the following wedge product equation:

$$a_1 \wedge a_2 \wedge a_3 \wedge a_4 \wedge \tilde{\mathbf{x}}_1 \wedge \ldots \wedge \tilde{\mathbf{x}}_m = 0. \qquad (3)$$

This wedge product constraint is referred as multilinear (in $\mathbf{x}_i$'s) constraints in computer vision literature and plays a very important role in motion recovery: since only images and the motion parameters are involved, it decouples the problem of motion recovery from the unknown structure. Note that this constraint is invariant under projective transformation, it is also the starting point of projective reconstruction (Triggs, 1996).

An alternative way of writing the wedge product $\wedge_{k=1}^4 a_k \wedge_{i=1}^m \tilde{\mathbf{x}}_i$ is in the standard form:

$$\sum_{1 \le i_1 < \cdots < i_{m+4} \le 3m} f_{i_1, \ldots, i_{m+4}} (e_{i_1} \wedge \ldots \wedge e_{i_{m+4}})$$

where $\{e_l\}_{l=1}^{3m}$ is the standard basis for $\mathbb{R}^{3m}$ and coefficients f's depend on a_k's and $\mathbf{x}_i$'s. In general, there are $\begin{pmatrix} 3m \\ m+4 \end{pmatrix}$ many ways how to choose $m+4$ ascending indexes from $\{1, \ldots, 3m\}$, which correspond to all the linearly independent terms. By the wedge product constraint (3), all the coefficients f's have to be zero. This gives the same number of homogeneous constraints of degree $m+4$ in terms of the entries of a_k's and $\tilde{\mathbf{x}}_i$'s. However, since many of the entries of $\tilde{\mathbf{x}}_i$'s are zeros, most of the homogeneous constraints will be trivial or reducible. In fact, nontrivial constraints imposed by equation (3) are either *bilinear*, *trilinear* or *quadrilinear* in the entries of $\mathbf{x}_i$'s (Triggs, 1995; Ma *et al.*, 1998*a*). If we view these f's as polynomials of entries of $\mathbf{x}_i$'s, then the coefficients [5] of these polynomials are functions (also polynomials) of entries of a_k's. Information about the camera motion g_i's and calibration A_i's are then encoded in these coefficients. It is then natural to ask what are the dependency among these coefficients of the bilinear, trilinear or quadrilinear constraints. Or do coefficients of trilinear and quadrilinear constraints necessarily give extra information about the motion matrix $\mathbf{A}$ than those of the bilinear constraints? The following theorem from (Ma *et al.*, 1998*c*) gives the answer:

Theorem 2. (Multilinear coefficients). For m images of a point, the coefficients of all bilinear constraints uniquely determine coefficients of trilinear and quadrilinear coefficients, given that the camera centers are not collinear.

In principle, if the camera is calibrated, the camera motion g_i's can then be estimated from these multilinear constraints, and knowing the camera motion, the structure, *i.e.* the unknown scales λ_i's, can be further reconstructed. See (Ma *et al.*, 1998*a*) for a more detailed account of existing algorithms for Euclidean motion and structure reconstruction.

[5] For example, for polynomials of a single indeterminant, the coefficients of $a_n x^n + \ldots + a_1 x + a_0$ are just $a_0, \ldots, a_n$.

4. EPIPOLAR GEOMETRY AND CAMERA SELF-CALIBRATION

In this section, we study how to recover motion and structure when the camera is not calibrated, *i.e.* the matrix A in (2) is unknown. In other words, can we recover the calibration directly from the image $\mathbf{x}$ without knowing the camera motion and scene structure? This is the so-called *camera self-calibration problem*, first studied by (Maybank and Faugeras, 1992) from a projective and algebraic geometric viewpoint. However, we here show that, in our framework, it is not necessary to introduce projective geometry, and instead self-calibration can be solved more elegantly using only linear algebra and a little bit differential geometry.

For demonstration, we here only consider the case that the calibration matrix A is time-invariant, *i.e.* all the A_i's in the motion matrix $\mathbf{A}$ are the same. [6] Recall that the unknown matrix A models the unknown linear transformation from $\mathbb{RP}^2$ to $\mathbb{RP}^2$ and can be assumed to be an element in $SL(3)$ (3 by 3 matrices of determinant $+1$). However, it can be shown that, without knowing the camera motion and scene structure, the camera self-calibration problem is geometrically equivalent to the problem of recovering an unknown metric given by the symmetric matrix $S = A^{-T}A^{-1}$ in a Euclidean space with a *calibrated* camera (Ma *et al.*, 1998*b*). [7]

Consider $m = 2$ for the multilinear constraint (3) and suppose that (R, p) is the relative motion between the two camera positions. It happens that the bilinear constraints given by (3) coincide with the so-called *epipolar constraint* of the form:

$$\mathbf{x}_1^T F \mathbf{x}_2 = 0 \qquad (4)$$

where $F \in \mathbb{R}^{3\times 3}$ is the so-called *fundamental matrix*:

$$F = A^{-T} R^T \widehat{p} A^{-1}$$

where $\widehat{p}$ is the skew-symmetric matrix associated to the vector p such that $\widehat{p}u = p \times u$ for any $u \in \mathbb{R}^3$. Since $A \in SL(3)$, we have $A^{-T}\widehat{p}A^{-1} = \widehat{Ap}$. Let $p' = Ap$ and we have $F = A^{-T}R^T A^T \widehat{p'}$ (then p' is the right kernel of the fundamental matrix F). It is direct to check that $S^{-1} = AA^T$ satisfies the equation:

[6] Analysis for the time-varying case is quite similar and one may refer to (Ma *et al.*, 1998*b*) for the detail.
[7] Hence the matrix A can only be recovered as an element in the quotient space $SL(3)/SO(3)$.

$$F^T S^{-1} F = \widehat{p'}^T S^{-1} \widehat{p'}. \qquad (5)$$

We call it the *normalized matrix Kruppa equation.*[8] Since the epipolar constraint (4) is homogeneous in F, F can only be estimated up to a scale. So in general, it is of the form $F = \lambda A^{-T} R^T A^T \widehat{p'}$ for some unknown scalar $\lambda \in \mathbb{R}$. Without loss of generality, we may always assume p' is of unit length. Then the normalized matrix Kruppa equation is modified to:

$$F^T S^{-1} F = \lambda^2 \widehat{p'}^T S^{-1} \widehat{p'}. \qquad (6)$$

We call this equation simply as the *matrix Kruppa equation.* Each matrix Kruppa equation gives two algebraically independent constraints on S hence in general at least three such equations are needed for solving S (Maybank and Faugeras, 1992). The Kruppa equation (6) does not show up without a reason – it is in fact directly related to invariants of the isometry group of the Euclidean space with metric S (Ma *et al.*, 1998*b*).

We are interested in under what conditions the Kruppa equations may have a unique solution. Given a fundamental matrix $F = A^{-T} R^T A^T \widehat{p'}$, the normalized matrix Kruppa equation (5) can be rewritten in the following way:

$$\widehat{p'}^T (S^{-1} - ARA^{-1} S^{-1} A^{-T} R^T A^T) \widehat{p'} = 0. \quad (7)$$

According to this form, if we define $C = A^{-T} R^T A^T$, a linear map $\sigma : \mathbb{R}^{3\times3} \to \mathbb{R}^{3\times3}$ as:

$$\sigma : X \mapsto X - C^T X C, \quad X \in \mathbb{R}^{3\times3},$$

and a linear map $\tau : \mathbb{R}^{3\times3} \to \mathbb{R}^{3\times3}$ as:

$$\tau : Y \mapsto \widehat{p'}^T Y \widehat{p'}, \quad Y \in \mathbb{R}^{3\times3},$$

then the solution S^{-1} of the equation (7) is exactly the (symmetric real) kernel of the composition map:

$$\tau \circ \sigma : \quad \mathbb{R}^{3\times3} \xrightarrow{\sigma} \mathbb{R}^{3\times3} \xrightarrow{\tau} \mathbb{R}^{3\times3}.$$

This interpretation of the Kruppa equation clearly decomposes effects of the rotational and translational parts of the motion: if there is no translation *i.e.* $p = 0$, then there is no map τ; if the translation is non-zero, the kernel is enlarged by composing the map τ. In general, the symmetric real kernel of the composition map $\tau \circ \sigma$ is of three dimension – while it can be shown that the kernel of σ is only of two dimension if R is not the identity (Ma *et al.*, 1998*b*).

Theorem 3. Given matrices $C_i = AR_iA^{-1}$, $i = 1, \ldots, m$ with $R_i \neq I$, $R \in SO(3)$, and given the real right (hence left) eigenvectors $u_i \in \mathbb{R}^3$ of R_i (*i.e.* the principal axis of the rotation matrix R_i), the symmetric real solution of equations $X - C_i^T X C_i = 0, i = 1, \ldots, m$ is unique if and only if at least two of the m principal axes $u_i, i = 1, \ldots, m$ are linearly independent.

This theorem says that for a set of Kruppa equations to have a unique solution, two of the rotation axes have to be linearly independent – a Lie theoretic explanation for this is that two is the minimal number of generators needed to generate the Lie algebra $so(3)$ of the Lie group $SO(3)$. For a more detailed analysis of the conditions of solving Kruppa equations, one may refer to (Ma *et al.*, 1998*b*), which also gives a more detailed account of the relation between the matrix Kruppa equation and the normalized one.

5. MOTION, STRUCTURE AND CALIBRATION UP TO SUBGROUPS

From the previous section, we know that the camera calibration can be uniquely recovered only if the camera motion satisfies certain conditions. We then should ask, if such conditions are not satisfied, to what extent we can reconstruct motion, structure and calibration? A natural approach to answer this question is to compute the corresponding generic ambiguities in reconstruction when the motion of the camera is restricted to a subgroup of $SE(3)$. A detailed study is provided in (Ma *et al.*, 1998*c*) where generic ambiguities associated to all Lie subgroups of $SE(3)$ are clearly studied and listed. It turns out that these ambiguities can be expressed as certain subgroups acting on the parameter space (of motion, structure and calibration). For example, if the camera motion is restricted to $SE(2) \subset SE(3)$, a planar motion, the corresponding ambiguity can expressed as a one parameter group, which essentially says that the relative scale along the direction orthogonal to the plane cannot be recovered (Ma *et al.*, 1998*c*).

As an application of these results, we can ask the following question: if the goal of the reconstruction is to produce a new two-dimensional image of the same scene structure from a different viewpoint, how and where can we generate "valid" images of the scene. Or in other words, in the presence of these reconstruction ambiguities, what are the proper camera positions allowing to generate images such that the ambiguities do not matter at all? This is so-called *reprojection problem* and has a lot of impact in applications such as video post-processing, new view synthesizing and vision-based navigation. With respect to different reconstruction ambiguities arising from different restricted camera motions, such proper camera positions can also be expressed as subgroups of

[8] Named after a mathematician named Kruppa who discovered a scalar version of this equation in 1913.

$SE(3)$, and a complete list has been given in (Ma *et al.*, 1998*c*).

6. MULTIVIEW GEOMETRY AND ITS APPLICATIONS

So far, we have outlined the main results in multiview geometry (of a vision system). Our interest is to apply these results to the control of mobile robot with an on-board vision system, *i.e.* to use the vision system as a sensor in the feedback control loop. Being able to provide relative orientation and velocities, such an on-board vision system certainly has its advantage over other motion sensors such as a gyroscope, accelerometer or even GPS.

Now consider a camera attached to a mobile robot, whose dynamics can be described by:

$$\begin{cases} \dot{g} = g\xi \\ \dot{\xi} = f(u) \end{cases}$$

where $g \in G$ and $\xi \in \mathfrak{g}$, the Lie algebra of G, and $f(u)$ is the force generated by the input u. Here, g describes the motion of the robot (or camera) and ξ is the velocity. The motion estimation scheme we studied above can therefore serves as an state observer to this system and its outputs, the motion g and velocity ξ, can then be used in various feedback controllers.

However, since motion estimation is not always a well-posed problem, people in practice prefer to design control laws using measurements directly from the image. This is the so called *visual servoing* approach. For instance, for tasks such as a car tracking ground curves, or an aircraft following terrain or landing autonomously, the control problems can be studied by lifting the curve or terrain dynamics onto the image plane. By studying the dynamics of the curve or surface in the image, one may design control laws using measurements directly from the image and bypass certain estimation issues (Ma *et al.*, to appear).

We have also been working on a graphical simulation of an autonomous aerial robot for the evaluation of control schemes and vision algorithms. The graphical simulation allows us to visualize the motion of the robot and evaluate the performance of the vision system's structure and motion recovery algorithms by directly comparing the estimates with the known information about the virtual environment.

Another application of vision is that of reconstructing a virtual environment from a sequence of images. We have shown that the structure of the environment can be recovered using vision; we can use this information to recreate a virtual environment. The potential applications of this technology are enormous and could range from virtual reality games to multimedia teleconferencing over the internet.

7. REFERENCES

Faugeras, O. (1993). *Three-Dimensional Computer Vision, a Geometric Viewpoint.* MIT Press.

Kobayashi, S. and K. Nomizu (1996). *Foundations of Differential Geometry.* Vol. I,II. John Wiley & Sons, Inc.

Koo, John, David Shim, Omid Shakernia, Bruno Sinopoli, Yi Ma, Frank Hoffmann and Shankar Sastry (1998). Hierarchical hybrid system design on Berkeley unmanned autonomous aerial vehicle.. *International Aerial Robotics Competition.*

Kosecka, J., R. Blasi, C.J. Taylor and J. Malik. (1997). Vision-based lateral control of vehicles. In: *Proc. Intelligent Transportation Systems Conference.* Boston.

Ma, Yi and Shankar Sastry (1998). Vision theory in space of constant curvature. *UC Berkeley Memorandum No. UCB/ERL M98/36.*

Ma, Yi, Jana Kosecka and Shankar Sastry (1998*a*). Euclidean structure and motion from image sequences. *UC Berkeley Memorandum No. UCB/ERL M98/38.*

Ma, Yi, Jana Kosecka and Shankar Sastry (1998*b*). A mathematical theory of camera self-calibration. *UC Berkeley Memorandum No. UCB/ERL M98/64.*

Ma, Yi, Jana Kosecka and Shankar Sastry (to appear). Vision guided navigation of a nonholonomic mobile robot. *IEEE Transactions on Robotics and Automation.*

Ma, Yi, Stefano Soatto, Jana Kosecka and Shankar Sastry (1998*c*). Euclidean reconstruction and reprojection up to subgroups. *UC Berkeley Memorandum No. UCB/ERL M98/65.*

Maybank, S. and O. Faugeras (1992). A theory of self-calibration of a moving camera. *International Journal of Computer Vision* **8**(2), 123–152.

Murray, Richard M., Zexiang Li and Shankar Sastry (1993). *A Mathematical Introduction to Robotic Manipulation.* CRC Press.

Triggs, B. (1995). Matching constraints and the joint image. In: *Proceeding of Fifth International Conference on Computer Vision.* IEEE Comput. Soc. Press. Cambridge, MA, USA. pp. 338–43.

Triggs, B. (1996). Factorization methods for projective structure and motion. In: *Proceeding of 1996 Computer Society Conference on Computer Vision and Pattern Recognition.* IEEE Comput. Soc. Press. San Francisco, CA, USA. pp. 845–51.