THREE-VIEW CAMERA CALIBRATION USING GEOMETRIC ALGEBRA

A. Dell'Acqua *A. Sarti* *S. Tubaro*

Dip. di Elettronica e Informazione – Politecnico di Milano,
Piazza Leonardo Da Vinci 32, 20133 Milano, Italy

ABSTRACT

In a former work of ours [1], we proposed a new way to express and interpret the epipolar constraint using Geometric Algebra, and we derived from it a novel and efficient 2-view camera calibration technique. In this paper we extend this GA approach to the 3-view case. After expressing the trifocal constraint in terms of bivectors and trivectors, we provide an alternative geometric interpretation of the coefficients of the trifocal tensor. On the basis of that, we propose a novel solution for the simultaneous determination of the focal lengths of the cameras and the rigid motion between three views.

1. INTRODUCTION

The structure from motion (SfM) problem is traditionally formulated in terms of geometric constraints, which are usually expressed using tools of linear algebra [2]. A more effective approach to the SfM problem exploits rank conditions on a properly defined multi-view matrix [8]. With such a purely algebraic approach, however, it is quite difficult to visualize the constraints in a geometric space. It would thus be highly desirable to devise and develop SfM algorithms that allow us to devise and enforce complex constraints in a visual fashion without having to give up the effectiveness and the generality of the algebraic solutions available today.

Geometric (Clifford) algebra (GA) is a framework where geometry and algebra sinergically co-exist. Using GA the SfM problem can be nicely formulated in a very evocative fashion (see [3, 7, 6]). In a previous work of ours ([1]) we showed how GA can be used to compactly rewrite the epipolar constraint and to derive novel geometric insight on the matter. In particular, we used an alternative camera frame based on camera bivectors (which correspond to the camera axes), and showed that the fundamental matrix and the rotation matrix that link a pair of views correspond to the components at infinity and the finite components, respectively, of the frame's bivectors (axes). We also proposed a novel 2-view self-calibration technique based on this interpretation. In this paper we extend this method to the 3-view case. In fact we will interpret the trifocal tensor and the camera motion in terms of finite components and components at infinity of intersections between camera planes (trivectors). Based on this method, we propose a 3-view calibration method that improves the performance of our 2-view calibration approach [1], which can be easily embedded in a camera tracking SW, as shown in the last Section.

2. THE PROJECTIVE SPACE IN GA

A generic point $\mathbf{p}$ of the projective space $\mathbb{P}^3$ can be written in homogeneous form as $\mathbf{p} = a_1\mathbf{e}_1 + a_2\mathbf{e}_2 + a_3\mathbf{e}_3 + \mathbf{e}_4$, where $\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3, \mathbf{e}_4$ form a basis of $\mathbb{P}^3$. The line $\mathbf{l}$ passing through a given pair of points $\mathbf{p}_1$ and $\mathbf{p}_2$ can be expressed as a bivector of the form $\mathbf{l} = \mathbf{p}_1 \wedge \mathbf{p}_2$, where the wedge operator denotes the *outer product* between vectors. Similarly, the plane passing through the three points $\mathbf{p}_1$, $\mathbf{p}_2$ and $\mathbf{p}_3$ can be written as the grade-3 vector (trivector) $\pi = \mathbf{p}_1 \wedge \mathbf{p}_2 \wedge \mathbf{p}_3$. Taking the outer product of a plane with a point that does not lie on the plane, we obtain a scaled version of the entire space ($\mathbf{I}_4$), called *pseudoscalar*. A generic multi-vector can be expressed as the linear combination of the basis element of the same grade. So a line $\mathbf{l}$ can be written as the linear combination of the basis bivectors as

$$\mathbf{l} = \alpha_1\mathbf{b}_1 + \alpha_2\mathbf{b}_2 + \alpha_3\mathbf{b}_3 + \beta_1\hat{\mathbf{b}}_1 + \beta_2\hat{\mathbf{b}}_2 + \beta_3\hat{\mathbf{b}}_3 \,, \quad (1)$$

where $\mathbf{b}_1 = \mathbf{e}_2\wedge\mathbf{e}_3$, $\mathbf{b}_2 = \mathbf{e}_3\wedge\mathbf{e}_1$, $\mathbf{b}_3 = \mathbf{e}_1\wedge\mathbf{e}_2$ are bivector on the plane at infinity π_∞ (component 'at infinity'), while $\hat{\mathbf{b}}_1 = \mathbf{e}_4 \wedge \mathbf{e}_1$, $\hat{\mathbf{b}}_2 = \mathbf{e}_4 \wedge \mathbf{e}_2$ and $\hat{\mathbf{b}}_3 = \mathbf{e}_4 \wedge \mathbf{e}_3$ pass through the origin ('finite' component). Similarly, a plane can be expressed as a linear combination of the basis trivectors of $\mathbb{P}^3$, which are the three planes through the origin ($\mathbf{t}_1 = \mathbf{e}_2 \wedge \mathbf{e}_3 \wedge \mathbf{e}_4$, $\mathbf{t}_2 = \mathbf{e}_3 \wedge \mathbf{e}_1 \wedge \mathbf{e}_4$ and $\mathbf{t}_3 = \mathbf{e}_1 \wedge \mathbf{e}_2 \wedge \mathbf{e}_4$) and the plane at infinity π_∞ ($\mathbf{t}_4 = \mathbf{e}_3 \wedge \mathbf{e}_1 \wedge \mathbf{e}_2$).

3. THREE-VIEW ANALYSIS IN GA

Let us assume, for the moment, that the focal lengths of the three cameras are equal to 1. As the trifocal constraint will be derived from a correspondence condition that involves line features on the image plane, it is more convenient to parametrize the cameras through its camera planes (the planes spanned by its axes).

Given a camera with center $\mathbf{c}$ and directions of the axes

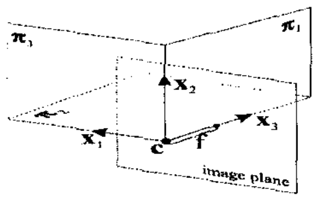

Fig. 1. A camera is completely defined by its axes and center, or by the three camera planes, and its focal length.

$\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3$ its camera planes are defined as

$$\pi_1 = \mathbf{c} \wedge \mathbf{x}_2 \wedge \mathbf{x}_3,\ \pi_2 = \mathbf{c} \wedge \mathbf{x}_3 \wedge \mathbf{x}_1,\ \pi_3 = \mathbf{c} \wedge \mathbf{x}_1 \wedge \mathbf{x}_2.$$

Using the *meet* operator ($\vee$), that represents in GA the intersection of subspaces, the camera axes can be expressed as

$$\mathbf{x}_1 = \pi_2 \vee \pi_3,\ \mathbf{x}_2 = \pi_3 \vee \pi_1,\ \mathbf{x}_3 = \pi_1 \vee \pi_2,$$

and the optical center is $\mathbf{c} = \pi_1 \vee \pi_2 \vee \pi_3$.

A camera plane parametrization simplifies the specification of line features on the image plane. In fact, the back-projection π_l of a line $\mathbf{l} = \begin{bmatrix} l_1 & l_2 & l_3 \end{bmatrix}^T$ on the image plane can be written as a linear combination

$$\pi_l = l_1\pi_1 + l_2\pi_2 + l_3\pi_3 \tag{2}$$

of the camera planes π_1, π_2 and π_3.

3.1. The Trifocal constraint

Let us consider a line $\mathbf{L}$ in 3-space, which projects onto the three image lines $\mathbf{l}^{(1)}$, $\mathbf{l}^{(2)}$ and $\mathbf{l}^{(3)}$. The back-projections of such lines

$$\pi_l^{(i)} = l_1^{(i)}\pi_1^{(i)} + l_2^{(i)}\pi_2^{(i)} + l_3^{(i)}\pi_3^{(i)} = \sum_{r=1}^{3} l_r^{(i)}\pi_r^{(i)} \quad i = 1,2,3 \quad ,$$

are bound to intersect in $\mathbf{L}$.

The line $\mathbf{L}$ in 3-space can also be written in terms of the *meet* operator as

$$\mathbf{L} = \pi_l^{(2)} \vee \pi_l^{(3)} = \sum_{s=1}^{3}\sum_{t=1}^{3} l_s^{(2)} l_t^{(3)} (\pi_s^{(2)} \vee \pi_t^{(3)}) \quad .$$

(the meet operator is distributive with respect to the sum).

If the lines $\mathbf{l}^{(1)}$, $\mathbf{l}^{(2)}$ and $\mathbf{l}^{(3)}$ correspond to the same line $\mathbf{L}$, then the outer product of the intersection $(\pi_l^{(2)} \vee \pi_l^{(3)})$ and the center of the first camera $\mathbf{c}^{(1)}$ must be equal to the back-projection plane $\pi_l^{(1)}$. We can thus write the trifocal constraint as

$$\pi_l^{(1)} = \mathbf{c}^{(1)} \wedge (\pi_l^{(2)} \vee \pi_l^{(3)}) \quad . \tag{3}$$

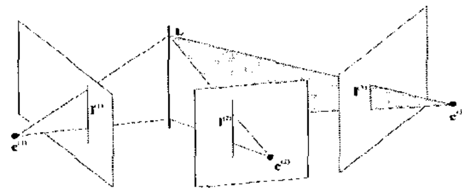

Fig. 2. Three image lines are projections of the same line in the space (i.e. satisfie the trifocal constraint) if and only if the corresponding back-projection planes intersect in a single line in the 3-space.

We can simplify this expression by computing the outer product between $\pi_l^{(1)}$ and the direction $\mathbf{x}_j^{(1)}$, $j = 1,2,3$, of each one of the axes of the first camera. Choosing for example $\mathbf{x}_1^{(1)}$, the left member becomes :

$$\pi_l^{(1)} \wedge \mathbf{x}_1^{(1)} = l_1^{(1)}(\mathbf{c}^{(1)} \wedge \mathbf{x}_2^{(1)} \wedge \mathbf{x}_3^{(1)} \wedge \mathbf{x}_1^{(1)}) = l_1^{(1)}\mathbf{I}_4 \tag{4}$$

Therefore eq. (4) becomes

$$\begin{aligned} l_1^{(1)}\mathbf{I}_4 &= \mathbf{c}^{(1)} \wedge \mathbf{x}_1^{(1)} \wedge (\pi_l^{(2)} \vee \pi_l^{(3)}) = \\ &= \sum_{s=1}^{3}\sum_{t=1}^{3} l_s^{(2)} l_t^{(3)} (\mathbf{c}^{(1)} \wedge \mathbf{x}_1^{(1)} \wedge (\pi_s^{(2)} \vee \pi_t^{(3)})) \quad . \end{aligned}$$

As all terms $(\mathbf{c}^{(1)} \wedge \mathbf{x}_1^{(1)} \wedge (\pi_s^{(2)} \vee \pi_t^{(3)}))$ are pseudoscalars, they are all proportional to $\mathbf{I}_4$. The substitutions $(\mathbf{c}^{(1)} \wedge \mathbf{x}_1^{(1)} \wedge (\pi_s^{(2)} \vee \pi_t^{(3)})) = \tau_{1st}\mathbf{I}_4$ lead to

$$l_1^{(1)} = \sum_{s=1}^{3}\sum_{t=1}^{3} l_s^{(2)} l_t^{(3)} \tau_{1st} \quad .$$

Similar expressions can be derived for $l_2^{(1)}$ and $l_3^{(1)}$. We have thus re-derived the classic formulation of the trifocal constraint, where the elements τ_{rst} are exactly the elements of the Hartley tensor [2].

3.2. Geometric interpretation of the Trifocal tensor

Exactly as seen in [1], we will now show that the elements of the trifocal tensor can be geometrically interpreted in terms of components at infinity of the camera parametrization. As seen above, the elements of the trifocal tensor can be written in terms of GA operators as

$$\tau_{rst}\mathbf{I}_4 = (\mathbf{c}^{(1)} \wedge \mathbf{x}_r^{(1)}) \wedge (\pi_s^{(2)} \vee \pi_t^{(3)}) \quad . \tag{5}$$

As in the bifocal case, we assume that the axes of the world coordinate frame are oriented like the axes of the first camera, and that the origin of the world frame is in the camera's optical center. Then eq. (5) becomes

$$\tau_{rst}\mathbf{I}_4 = (\mathbf{e}_4 \wedge \mathbf{e}_r) \wedge (\pi_s^{(2)} \vee \pi_t^{(3)}) \quad .$$

This expression, as shown in [1], is equivalent to $\tau_{rst} = (\pi_s^{(2)} \vee \pi_t^{(3)}) \cdot \mathbf{b}_r$: the 27 elements of the trifocal tensor τ_{rst} are, in fact, the coefficients of the component at infinity of the 9 lines $(\pi_s^{(2)} \vee \pi_t^{(3)})$. These lines are the intersections of the planes of the second camera and the planes of the third camera. Notice that in the bifocal case the unknown lines were the axes of the second camera. Such axes are obtained as mutual intersections between the 3 planes of the second camera.

Since the three lines of intersection between one camera plane and the three planes of the other camera are bound to be coplanar, they cannot be independent. This leads to nonlinear constraints among the trifocal tensor entries (see [4, 5]), which reduce to 18 the number of degrees of freedom of the trifocal tensor.

3.3. Retrieving the cameras

The trifocal tensor can be computed directly from line (or point) correspondences on the images. In order to compute position and orientation of the second and third camera we need to find a system of equations to retrieve the expression of the planes of the second and third camera from the trifocal tensor. The first camera is assumed to be in canonical position. The planes of the k-th camera can be written as

$$\pi_i^{(k)} = c_{i1}^{(k)}\mathbf{t}_1 + c_{i2}^{(k)}\mathbf{t}_2 + c_{i3}^{(k)}\mathbf{t}_3 + c_{i4}^{(k)}\mathbf{t}_4 \quad ,$$

where $c_{ij}^{(k)}$ are the unknowns that we must retrieve. Given two of these planes we can directly compute their *meet* as

$$\begin{aligned}(\pi_s^{(2)} \vee \pi_t^{(3)}) = \\ = (c_{s1}^{(2)}c_{t4}^{(3)} - c_{s4}^{(2)}c_{t1}^{(3)})e_{2,3} + (c_{s2}^{(2)}c_{t4}^{(3)} - c_{s4}^{(2)}c_{t2}^{(3)})e_{3,1} + \\ +(c_{s3}^{(2)}c_{t4}^{(3)} - c_{s4}^{(2)}c_{t3}^{(3)})e_{1,2} + (c_{s3}^{(2)}c_{t2}^{(3)} - c_{s2}^{(2)}c_{t3}^{(3)})e_{4,1} + \\ +(c_{s1}^{(2)}c_{t3}^{(3)} - c_{s3}^{(2)}c_{t1}^{(3)})e_{4,2} + (c_{s2}^{(2)}c_{t1}^{(3)} - c_{s1}^{(2)}c_{t2}^{(3)})e_{4,3} \quad .\end{aligned}$$

The elements of the trifocal tensor correspond to the coefficients of the component at infinity of the lines $\pi_s^{(2)} \vee \pi_t^{(3)}$, $r, s = 1, 2, 3$. This provides us with a set of 27 equations of the form

$$\tau_{rst} = c_{sr}^{(2)}c_{t4}^{(3)} - c_{s4}^{(2)}c_{tr}^{(3)} \quad , \tag{6}$$

which are 27 nonlinear constraints on the 24 unknowns.

We now need a second set of geometric constraints to impose the orthogonality between the three planes of one camera. We know that the orientation of a subspace of $\mathbb{P}^3$ can be obtained by computing its intersection with the plane at infinity π_∞. The meet of a plane π with π_∞ is

$$\pi \vee \pi_\infty = \sum_{j=1}^{4} c_j \mathbf{t}_j \vee \pi_\infty = \sum_{j=1}^{3} c_j \mathbf{b}_j \quad ,$$

which is a line on the plane at infinity [1]. Notice that the vector of the coefficients $\begin{bmatrix} c_1 & c_2 & c_3 \end{bmatrix}^T$ of the intersection line is also the direction of the vector normal to the plane.

[1] given two subspaces $\mathbf{A}$ and $\mathbf{B}$ whose join is the pseudoscalar, their meet can be rewritten as $\mathbf{A}^* \cdot \mathbf{B}$, where $\mathbf{A}^*$ is the dual subspace of $\mathbf{A}$ [7], therefore we have $\pi_\infty \vee \pi = \pi_\infty^* \cdot \pi = \pi \cdot (-\mathbf{e}_4)$)

It can be easily shown that, given the three orthogonal planes $\pi_{i=1..3}^{(k)}$ of the k-th camera, the matrix that rotates the 'finite' basis $\mathbf{t}_1$, $\mathbf{t}_2$ and $\mathbf{t}_3$ in $\pi_{i=1..3}^{(k)}$ is

$$\mathbf{R}^{(k)} = \begin{bmatrix} c_{11}^{(k)} & c_{21}^{(k)} & c_{31}^{(k)} \\ c_{12}^{(k)} & c_{22}^{(k)} & c_{32}^{(k)} \\ c_{13}^{(k)} & c_{23}^{(k)} & c_{33}^{(k)} \end{bmatrix} \quad .$$

whose columns are exactly the first three coefficients of the planes $\pi_{i=1..3}^{(k)}$. Representing $\mathbf{R}^{(k)}$ with a quaternion, exactly as done in the bifocal case, eq. (6) can be rewritten with a reduced number of unknowns (from nine to four), thus leading to a more efficient and robust parametrization of the rotation. In addition, the orthogonality constraint on the elements of $\mathbf{R}^{(k)}$ becomes a simple normalization constraint on the equivalent quaternion.

3.4. Calibration in the trifocal case

If we remove the assumption that the focal lengths are equal to one, the equation that describes a back-projected line as a linear combination of camera planes, becomes

$$\pi_l = l_1 f \pi_1 + l_2 f \pi_2 + l_3 \pi_3 \quad . \tag{7}$$

Let $f^{(1)}$, $f^{(2)}$ and $f^{(3)}$ be the focal lenghts of the first, second and third camera, respectively. Following the same steps as in the 'normalised' case , we can start from eq. (7) and prove that

$$\begin{aligned}\hat{\tau}_1 &= \begin{bmatrix} (\frac{f^{(2)}f^{(3)}}{f^{(1)}})\tau_{111} & (\frac{f^{(2)}f^{(3)}}{f^{(1)}})\tau_{112} & (\frac{f^{(2)}}{f^{(1)}})\tau_{113} \\ (\frac{f^{(2)}f^{(3)}}{f^{(1)}})\tau_{121} & (\frac{f^{(2)}f^{(3)}}{f^{(1)}})\tau_{122} & (\frac{f^{(2)}}{f^{(1)}})\tau_{123} \\ (\frac{f^{(3)}}{f^{(1)}})\tau_{131} & (\frac{f^{(3)}}{f^{(1)}})\tau_{132} & (\frac{1}{f^{(1)}})\tau_{133} \end{bmatrix} \\ \hat{\tau}_2 &= \begin{bmatrix} (\frac{f^{(2)}f^{(3)}}{f^{(1)}})\tau_{211} & (\frac{f^{(2)}f^{(3)}}{f^{(1)}})\tau_{212} & (\frac{f^{(2)}}{f^{(1)}})\tau_{213} \\ (\frac{f^{(2)}f^{(3)}}{f^{(1)}})\tau_{221} & (\frac{f^{(2)}f^{(3)}}{f^{(1)}})\tau_{222} & (\frac{f^{(2)}}{f^{(1)}})\tau_{223} \\ (\frac{f^{(3)}}{f^{(1)}})\tau_{231} & (\frac{f^{(3)}}{f^{(1)}})\tau_{232} & (\frac{1}{f^{(1)}})\tau_{233} \end{bmatrix} \\ \hat{\tau}_3 &= \begin{bmatrix} (f^{(2)}f^{(3)})\tau_{311} & (f^{(2)}f^{(3)})\tau_{312} & f^{(2)}\tau_{313} \\ (f^{(2)}f^{(3)})\tau_{321} & (f^{(2)}f^{(3)})\tau_{322} & f^{(2)}\tau_{323} \\ f^{(3)}\tau_{331} & f^{(3)}\tau_{332} & \tau_{333} \end{bmatrix} \quad ,\end{aligned} \tag{8}$$

where $\hat{\tau}_i$ $i = 1..3$ are 'slice' matrices of the trifocal tensor in the case of unknown focal lenghts, while τ_{rst} are the element of the trifocal tensor in the 'essential' (normalised) case. As eq. (6) holds true also in the case of unknown focal lenghts, the variables $f^{(1)}$, $f^{(2)}$ and $f^{(3)}$ can be simply added as unknowns to the systems of equations (6), which provide enough constraints to retrieve position and orientation of the second and third camera, as well as the focal lenghts.

4. SIMULATION RESULTS

In [1] we proved that our 2-view calibration method performs comparably with state-of-art methods in the literature. We conducted a series of experiments on noisy image

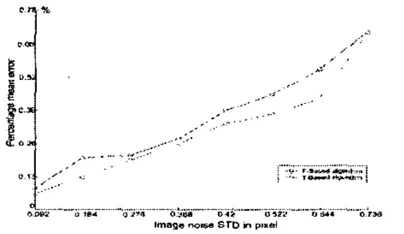

Fig. 3. Average relative error on the focal length estimation in the case of three cameras with fixed focal length. The calibration method based on three views is more accurate and stable than in the two-view case for all noise levels.

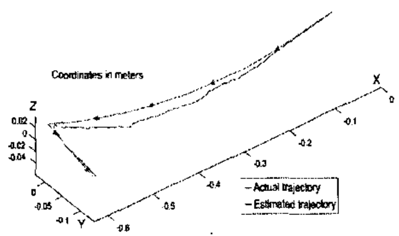

Fig. 4. Actual and estimated camera trajectories of the same sequence.

coordinates of point clouds with the goal of comparing the performance of our 3-view method with the 2-view one and assess the impact of the third view on the global accuracy. Fig. 3 shows the average relative errors in the estimation of a fixed focal length, obtained when using the two-view and the three-view methods. Notice,however, that the fundamental matrices of the 2-view estimator are here robustly computed through using the trifocal tensor. This means that what we measure now is what we gain from computing focal lengths directly from the trifocal tensor. The calibration method based on 3 views, in spite of the increasing number of parameters to be simultaneously recovered, turns to be more accurate and stable than the 2-view based algorithm for all noise levels. In both the cases the system of equations to compute camera orientations and positions was solved numerically using a Newton-Raphson algorithm. In order to test the effectiveness of the proposed calibration approach in real applications, we also implemented a simple camera-tracker based on it. Our 3-view algorithm is here used to calibrate single triplets of frames, in order to obtain partial reconstructions, which are then merged into a global one. To do this, the value of the focal length was set to the average value computed from the triplets that showed the lowest back-projection error, and then the 3-view algorithm was re-run while holding the focal length fixed. The accuracy evaluation of the camera tracker was conducted on sequences with known motion. In order to do so, we mounted a camera on a high-precision mechanical arm whose motion was controlled by a PC. This way we could compare the estimated trajectory with the actual (programmed) camera motion. The acquired scene contained two objects (a teddy bear and a puppet, each about 30 cm tall) with some natural texturing on them. The mechanical arm kept the camera at a distance of approximately 30-40cm from the objects. A visual comparison between estimated and actual trajectory can be seen in Fig. 4 on the right. As we can see, the estimated camera trajectory turned out to be very close to the actual one.

5. CONCLUSIONS

In this paper we proposed a novel geometric interpretation of the trifocal tensor in terms of bivectors and trivectors in GA. From this parameterization we derived a procedure for computing unknown focal lengths as well as camera positions and orientations, without introducing projective ambiguities.

6. REFERENCES

[1] C. Defferara, F. Negroni, A. Sarti, S. Tubaro, "New Perspectives on Camera Calibration Using Geometric Algebra". *Proc. ICIP 2002*, New York, Sept. 2002.

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

[3] E. Baryo-Corrachano, J. Lasenby, G. Sommer, "Geometric Algebra: A Framework for Computing Point and Line Correspondances and Projective Structure Using *n* Uncalibrated Cameras". *ICPR-96*, Vienna, Aug. 1996.

[4] C.B.U. Perwass, J. Lasenby, "A Geometric Derivation of the Trifocal Tensor and its Constraints", Cambridge University Engineering Department, Technical Report CUED/F - INFENG/TR.331, Oct. 1998.

[5] O. Faugeras, T. Papadopoulo, "A Nonlinear Method for Estimating the Projective Geometry of 3 Views". *ICCV-98*, pp. 477-484, 1998.

[6] O. Faugeras, B. Mourrain, "On the Geometry and Algebra of the Point and Line Correspondances between N Images". *ICCV-95*, pp. 951-962, 1995.

[7] C.B.U. Perwass, *Application of Geometric Algebra in Computer Vision*, Ph.D. Thesis, University of Cambridge, 2000.

[8] Y. Ma, K. Huang, R. Vidal, J. Kosecka, S. Sastry, *Rank conditions on the multiple view matrix*, Kluwer Academic Publishers, 2001.