Euclidean and Affine Structure/Motion for Uncalibrated Cameras from Affine Shape and Subsidiary Information *

Gunnar Sparr

Dept. of Mathematics, Lund University/LTH,
P.O. Box 118, S-22100 Lund, Sweden
gunnar@maths.lth.se

Abstract. The paper deals with the structure–motion problem for uncalibrated cameras, in the case that subsidiary information is available, consisting e.g. in known coplanarities or parallelities among points in the scene, or known positions of some focal points (hand-eye calibration). Despite unknown camera calibrations, it is shown that in many instances the subsidiary information makes affine or even Euclidean reconstruction possible. A parametrization by affine shape and depth is used, providing a simple framework for the incorporation of apriori knowledge, and enabling the development of iterative, rapidly converging algorithms. Any number of points in any number of images are used in a uniform way, with equal priority, and independently of coordinate representations. Moreover, occlusions are allowed.

1 Introduction

The structure and motion problem is central for computer vision, dealing with the analysis of a 3D scene by means of a sequence of 2D images. It is often studied by epipolar geometry and multilinear constraints, cf. [2], [3], [4], [5], [6], [7]. [10], [11], [19], [21]. The present paper uses an alternative approach, based on the notions of affine shape and depth, developed in a series of papers [12], [13], [14], [15], [16], [17], [18].

Depending on the apriori information available, the structure and motion problem can be treated on different levels. In the case of uncalibrated cameras it is well known that only projective reconstruction is possible, cf. [2], [14]. Working with point configurations, we here consider the case when some affine or Euclidean knowledge about the scene or the camera locations is available, e.g. a number of occurences like 'two lines are parallel' or 'a line is parallel to a plane', in which case affine reconstruction can be achieved. When having in addition some sort of Euclidean information, like a city map in the case of pictures of a city scene, this may be strengthened to Euclidean reconstruction. Another situation considered is when the relative placement of at least five focal points

* The work has been supported by the ESPRIT reactive LTR project 21914, CUMULI, and by the Swedish Council for Engineering Sciences (TFR), project 95-64-222.

are known, where is is shown that not only the projective but also the affine structure of the scene can be recovered. Again, having Euclidean information about the focal points, this can be strengthened to Euclidean reconstruction of the scene. The latter situation appears naturally in hand-eye calibration from pictures taken by a camera mounted on a moving robot arm, with registration of the motion parameters. It is also shown how to adjust a reconstruction to be consistent with coplanarity constraints for some set of points.

The notion of affine shape is well suited to handle these situations, theoretically as well as computationally. To gain robustness, numerical computations are based on a variational formulation, possible to exploit by linear algebraic methods. Data from any number of points in any number of images can be treated simultaneously, without preselection of reference points or images. In particular, there is no need to handle the numerically unstable situation of overdetermined systems of polynomial equations with uncertain coefficients.

The plan of the paper is as follows. In Section 2 a brief recapitulation of the notions of affine depth and shape and their use in single view geometry is given. Section 3 deals with multiple view geometry along the same lines. With this background, Section 4 presents algorithms for projective reconstruction. These are then extended in Section 5 to affine and Euclidean reconstruction in the case of subsidiary information. More details and discussions about the notions of affine shape and depth are given in a self-contained and independent Appendix.

2 Single view geometry by affine shape and depth

Let $\mathbf{A}^3$ denote the three-dimensional affine space, let Π' be a plane in $\mathbf{A}^3$, the *image plane*, and let Π be a subset of $\mathbf{A}^3$. By P_ϕ is meant the *perspective transformation* $\Pi \to \Pi'$ with *centre* ϕ. Here ϕ is allowed to be a point at infinity, in which case P_ϕ is a *parallel projection* in direction ϕ. Perspective transformations model the *pinhole camera*. If no metrical information is known or used, the camera is said to be *uncalibrated*.

The principal objects dealt with in this paper are *n-point configurations* $\mathcal{X}$, by which is meant ordered sets of points $\mathcal{X} = (X^1, \dots, X^n)$, where $X^k \in \mathbf{A}^3$, $k = 1, \dots, n$. Let $\rho_\mathcal{X}$ denote the *dimension* of $\mathcal{X}$, e.g. $\rho_\mathcal{X} = 3$ if $\mathcal{X}$ is a non-planar 3D-configuration. The set of n-point configurations of dimension ρ will be denoted $\mathcal{C}_{n,\rho}$.

In a series of papers [12], [13], [14], [15], the notions of *affine shape space* and *affine depth space* have been developed. The definitions and main properties are summarized below. For a somewhat more thorough presentation, see the accompanying appendix.

The following notation is used throughout the paper: If $\alpha = (\alpha_1, \dots, \alpha_n)$, $\xi = (\xi_1, \dots, \xi_n)$, let $\alpha\xi = (\alpha_1\xi_1, \dots, \alpha_n\xi_n)$, and let $\overline{\alpha} = (1/\alpha_1, \dots, 1/\alpha_n)$. Moreover, let $\Sigma_0 = \{\xi \in \mathbb{R}^n \mid \sum_1^n \xi_k = 0\}$.

- *Definition of affine shape and depth spaces.* Let x^k be the coordinate column vector of X^k with respect to an arbitrary affine basis, $k = 1, \dots, n$. Then the *affine shape space* and the *affine depth space* are defined by

$$s(\mathcal{X}) = \mathcal{N}\begin{bmatrix} 1 & 1 & \cdots & 1 \\ x^1 & x^2 & \cdots & x^n \end{bmatrix} \quad \text{and} \quad d(\mathcal{X}) = \mathcal{R}_{\text{row}}\begin{bmatrix} 1 & 1 & \cdots & 1 \\ x^1 & x^2 & \cdots & x^n \end{bmatrix} ,$$

respectively, where $\mathcal{N}$ stands for nullspace, and $\mathcal{R}_{\text{row}}$ for rowspace. (Cf. Definition A.1 of the Appendix.)

- *Affine invariancy.* There exists an affine transformation $A : \mathcal{X} \longrightarrow \mathcal{X}'$ if and only if $s(\mathcal{X}) \subset s(\mathcal{X}')$, or, equivalently, $d(\mathcal{X}') \subset d(\mathcal{X})$. If $\mathcal{X}$ is restricted to planar configurations, then the inclusions are in fact equalities, $s(\mathcal{X}) = s(\mathcal{X}')$ and $d(\mathcal{X}') = d(\mathcal{X})$, respectively. In this case we also write $\mathcal{X} \stackrel{s}{=} \mathcal{X}'$. (Cf. Theorem A.1.)
- *Dimension.* The dimensions of the linear spaces $s(\mathcal{X})$ and $d(\mathcal{X})$ are related to the dimension of the configuration by $\dim s(\mathcal{X}) = n - \rho_{\mathcal{X}} - 1$ and $\dim d(\mathcal{X}) = \rho_{\mathcal{X}} + 1$, respectively. (Cf. Theorem A.2.)
- *Shape and depth theorem.* There exists a perspective transformation P such that $P(\mathcal{X}) \stackrel{s}{=} \mathcal{Y}$ with depth α if and only if $\alpha s(\mathcal{X}) \subset s(\mathcal{Y})$, or, equivalently, $\alpha d(\mathcal{Y}) \subset d(\mathcal{X})$. If $\mathcal{X}$ is restricted to planar configurations, then the inclusions are replaced by $\alpha s(\mathcal{X}) = s(\mathcal{Y})$ and $\alpha d(\mathcal{Y}) = d(\mathcal{X})$, respectively. (Cf. Theorem B.1.)
- *Definition of S- and D-matrices.* By an S-matrix of $\mathcal{X}$ is meant a matrix having $s(\mathcal{X})$ as column space. By a D-matrix is meant a matrix having $d(\mathcal{X})$ as row space. Passage between different S-matrix representations is performed by multiplication from the right by a non-singular matrix. (Cf. Definition A.2.)
- *Focal point theorem.* If $\alpha s(\mathcal{X}) \subset s(\mathcal{Y})$, then there exists a projection $P_\phi : \mathcal{X} \longrightarrow \mathcal{Y}$ if and only if

$$\phi = \sum_{k=1}^{n} \overline{\alpha}_k \eta_k X^k \Big/ \sum_{k=1}^{n} \overline{\alpha}_k \eta_k ,$$

where $\eta \in s(\mathcal{Y}) \setminus \alpha s(\mathcal{X})$. The compound configuration $(\mathcal{X}, \phi)$ thus has an S-matrix

$$S_{(\mathcal{X},\phi)} = \begin{bmatrix} \operatorname{diag}(\overline{\alpha}) S_{\mathcal{Y}} \\ -\overline{\alpha}^T S_{\mathcal{Y}} \end{bmatrix} .$$

(Cf. Theorem B.2.)

3 Multiple views

3.1 Main theorem

Suppose that $\mathcal{Y}^1, \ldots, \mathcal{Y}^m \in \mathcal{C}_{n,2}$ are projective images of one and the same configuration $\mathcal{X} \in \mathcal{C}_{n,3}$. The shape and depth theorem implies that $\alpha^1 s(\mathcal{X}) \subset s(\mathcal{Y}^1), \ldots, \alpha^m s(\mathcal{X}) \subset s(\mathcal{Y}^m)$, or, equivalently,

$$s(\mathcal{X}) \subset \overline{\alpha}^1 s(\mathcal{Y}^1) \quad , \ldots , \quad s(\mathcal{X}) \subset \overline{\alpha}^m s(\mathcal{Y}^m) \ , \ \text{where } \alpha^1, \ldots, \alpha^m \in d(\mathcal{X}) \ .$$

From this the equivalence between the first two items in the following theorem follows. An analogous argument for the depth space yields the equivalence with the third item.

Theorem 1. Structure theorem. *Let $\mathcal{X} \in \mathcal{C}_{n,\rho}$. The following statements are equivalent:*

- $\mathcal{Y}^i \in \mathcal{C}_{n,\rho-1}$ *is a projective image of* $\mathcal{X}$ *with depth vector* α^i, $i = 1, \dots, m$, *where not all projections are flat,*
- $s(\mathcal{X}) = \overline{\alpha}^1 s(\mathcal{Y}^1) \bigcap \dots \bigcap \overline{\alpha}^m s(\mathcal{Y}^m)$,
- $d(\mathcal{X}) = \alpha^1 d(\mathcal{Y}^1) + \dots + \alpha^m d(\mathcal{Y}^m)$.

First note the ambiguity in the last two items, consisting in that they remain valid after multiplication with any β in $d(\mathcal{X})$, giving rise to a new consistent reconstruction, with shape space $\beta s(\mathcal{X})$. This is the shape-depth formulation of the well-known projective reconstruction ambiguity, cf. [14]. Also note that since $\dim d(\mathcal{X}) = 4$, the ambiguity is governed by four independent components of β.

To indicate the usage of the theorem, consider the equivalence of the first two items in the case $m = 2$. First normalize by multiplying by $\beta = \alpha^1$, then put $q^2 = \alpha^2/\alpha^1$, and let $\mathcal{X}_{\|}$ denote the corresponding reconstructed object configuration. Here q^2 is called *kinetic depth*, and the notation $\mathcal{X}_{\|}$ comes from the fact that $\mathcal{Y}^1$ is formed by a parallel projection. The condition to fulfill is

$$s(\mathcal{X}_{\|}) = s(\mathcal{Y}^1) \bigcap \overline{q}^2 s(\mathcal{Y}^2) \ . \tag{1}$$

To analyze this condition, choose S-matrices $S_{\mathcal{Y}^1}$, $S_{\mathcal{Y}^2}$, and form the compound matrix

$$W_q(\mathcal{Y}^1, \mathcal{Y}^2) = \left[S_{\mathcal{Y}^1} \mid \operatorname{diag}(\overline{q}^1) S_{\mathcal{Y}^2} \right] \ .$$

A dimension argument yields that a necessary condition for (1) to be fulfilled is that

$$\dim \mathcal{N} W_q(\mathcal{Y}^1, \mathcal{Y}^2) = n - \rho - 1 \quad \Longleftrightarrow \quad \operatorname{rank} W_q(\mathcal{Y}^1, \mathcal{Y}^2) = n - \rho + 1 \ .$$

One way to proceed is by forming polynomial equations from the vanishing of all subdeterminants of W_q of order $n - \rho$. For reasons that will be discussed in Section 3.3, we prefer another method, described in Section 4.1. However, once q^2 is determined, $\mathcal{X}_{\|}$ can be computed as the intersection space in (1), after which all other consistent reconstructions are obtained by multiplication with $\beta \in d(\mathcal{X}_{\|})$. As remarked above, this gives a four parameter family of solutions to the reconstruction problem.

3.2 The Chasles matrix

Next we combine Theorem 1 with the focal point theorem, to describe the interplay between structure and motion. By means of the S-matrices of the respective images and the depth vectors $\overline{\alpha}^1, \dots, \overline{\alpha}^m$, a compound matrix, called the *Chasles matrix*, is formed

$$C(\mathcal{Y}^1, \dots, \mathcal{Y}^m; \overline{\alpha}^1, \dots, \overline{\alpha}^m) = \begin{bmatrix} \operatorname{diag}(\overline{\alpha}^1) S_{\mathcal{Y}^1} & \cdots & \operatorname{diag}(\overline{\alpha}^m) S_{\mathcal{Y}^m} \\ -\overline{\alpha}^{1T} S_{\mathcal{Y}^1} & \cdots & 0 \\ 0 & \cdots & 0 \\ \cdot & \cdots & \cdot \\ 0 & \cdots & -\overline{\alpha}^{mT} S_{\mathcal{Y}^m} \end{bmatrix} . \tag{2}$$

Theorem 2. Structure and motion theorem. *Let $\mathcal{X} \in \mathcal{C}_{n,\rho}$. The following statements are equivalent.*

- *$\mathcal{Y}^i \in \mathcal{C}_{n,\rho-1}$ is a projective image of $\mathcal{X}$ with depth vector α^i, $i = 1, \dots, m$, where not all projections are flat,*
- *the Chasles matrix $C(\mathcal{Y}^1, \dots, \mathcal{Y}^m; \overline{\alpha}^1, \dots, \overline{\alpha}^m)$ is an S-matrix of the compound configuration $(\mathcal{X}, \phi^1, \dots, \phi^m)$.*

From this theorem it follows that a necessary and sufficient condition for geometric consistency is that the Chasles matrix has rank $m+n-4$. In particular, this means that when having fixed the locations of four of the $m+n$ points $X_1, \dots, X_n, \phi^1, \dots, \phi^m$, all the others are known too, as linear expressions in the four selected points. These expressions can be read out explicitely from the Chasles matrix, as illustrated by the following example.

Example 1. Let $\mathcal{Y}^1$ and $\mathcal{Y}^2$ be defined by their S-matrices

$$S_{\mathcal{Y}^1} = \begin{bmatrix} 8 & 8 \\ -4 & -1 \\ -1 & -4 \\ -3 & 0 \\ 0 & -3 \end{bmatrix} , \quad S_{\mathcal{Y}^2} = \begin{bmatrix} 1 & 5 \\ -1 & 2 \\ 1 & -4 \\ -1 & 0 \\ 0 & -3 \end{bmatrix} .$$

Then

$$W_q(\mathcal{Y}^1, \mathcal{Y}^2) = \begin{bmatrix} 8 & 8 & \overline{q}_1 & 5\overline{q}_1 \\ -4 & -1 & -\overline{q}_2 & 2\overline{q}_2 \\ -1 & -4 & \overline{q}_3 & -4\overline{q}_3 \\ -3 & 0 & -\overline{q}_4 & 0 \\ 0 & -3 & 0 & -3\overline{q}_5 \end{bmatrix} .$$

One verifies that if $\overline{q} = (3, 6, 9, 1, 2)$, then $\operatorname{rank} W_q = 3$. According to (2), a Chasles matrix is obtained by enlarging W_q with two rows, in such a way that all column sums vanish:

$$C(\mathcal{Y}^1, \mathcal{Y}^2, \mathbf{1}, q) = \begin{bmatrix} 8 & 8 & 3 & 15 \\ -4 & -1 & -6 & 12 \\ -1 & -4 & 9 & -36 \\ -3 & 0 & -1 & 0 \\ 0 & -3 & 0 & -6 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & -5 & 15 \end{bmatrix} .$$

Here the first five rows correspond to points of $\mathcal{X}$, while rows six and seven correspond to ϕ^1 and ϕ^2, respectively.

Another Chasles matrix is obtained by elimination of the ϕ^2-component in the fourth column:

$$C(\mathcal{Y}^1, \mathcal{Y}^2, \mathbf{1}, q) = \begin{bmatrix} 8 & 8 & 3 & 24 \\ -4 & -1 & -6 & -6 \\ -1 & -4 & 9 & -9 \\ -3 & 0 & -1 & -3 \\ 0 & -3 & 0 & -6 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & -5 & 0 \end{bmatrix} .$$

From the fourth and third columns of C we read out that

$$s(\mathcal{X}_{\|}) = \text{linear hull}\,(8, -2, -3, -1, -2) \quad \text{and} \quad \phi^2 = \frac{1}{5}(3X_1 - 6X_2 + 9X_3 - X_4) .$$

Moreover, from the first column it follows that P^1 is a parallel projection in the direction

$$-4\overline{X_1X_2} - \overline{X_1X_3} - 3\overline{X_1X_4} ,$$

which determines the point at infinity ϕ^1.

By this we have completely described one solution of the structure-motion problem. All other solutions are generated by letting $\overline{\alpha}^1$ run through $d(\mathcal{X}_{\|})$, i.e. the hyperplane $8\overline{\alpha}_1^1 - 2\overline{\alpha}_2^1 - 3\overline{\alpha}_3^1 - \overline{\alpha}_4^1 - 2\overline{\alpha}_5^1 = 0$, with four degrees of freedom. One example of such an $\overline{\alpha}^1$ is $(3, 3, 2, 6, 3)$. Then $\overline{\alpha}^2 = \overline{\alpha}^1\overline{q} = (9, 18, 18, 6, 6) \parallel (3, 6, 6, 2, 2)$, which gives the Chasles matrix

$$C(\mathcal{Y}^1, \mathcal{Y}^2, \alpha^1, \alpha^2) = \begin{bmatrix} 24 & 24 & 3 & 15 \\ -12 & -3 & -6 & 12 \\ -2 & -8 & 6 & -24 \\ -18 & 0 & -2 & 0 \\ 0 & -9 & 0 & -6 \\ 8 & -4 & 0 & 0 \\ 0 & 0 & -1 & 3 \end{bmatrix} .$$

After elimination of a ϕ^1-component of the first image, and a ϕ^2-component of the second, we obtain another Chasles matrix

$$C(\mathcal{Y}^1, \mathcal{Y}^2, \alpha^1, \alpha^2) = \begin{bmatrix} 24 & 72 & 3 & 24 \\ -12 & -18 & -6 & -6 \\ -2 & -18 & 6 & -6 \\ -18 & -18 & -2 & -6 \\ 0 & -18 & 0 & -6 \\ 8 & 0 & 0 & 0 \\ 0 & 0 & -1 & 0 \end{bmatrix} .$$

Now all characteristics of the structure–motion problem, the shape of $\mathcal{X}$ as well as the focal points, can be read out:

$$s(\mathcal{X}) = \text{linear hull}\,(4, -1, -1, -1, -1) ,$$
$$\phi^1 = -3X_1 + \tfrac{3}{2}X_2 + \tfrac{1}{4}X_3 + \tfrac{9}{2}X_4 ,$$
$$\phi^2 = 3X_1 - 6X_2 + 6X_3 - 2X_4 .$$

Note that columns two and four are parallel, and that both describe the shape of $\mathcal{X}$. This is what could be expected from the fact that the object configuration has not changed between the imaging instants. Also note that $\alpha^1 s(\mathcal{X}) = s(\mathcal{X}_{\|})$, in accordance with the discussion above.

3.3 Relation to fundamental matrices and multilinear forms

To fix the ideas, consider the case of 5-point configurations, and choose S-matrices so that

$$W_q(\mathcal{Y}^1, \mathcal{Y}^2) = \left[\begin{array}{cc|cc} \eta^1_{11} & \eta^1_{12} & \overline{q}_1\eta^2_{11} & \overline{q}_1\eta^2_{12} \\ \eta^1_{21} & \eta^1_{22} & \overline{q}_2\eta^2_{21} & \overline{q}_2\eta^2_{22} \\ \eta^1_{31} & \eta^1_{32} & \overline{q}_3\eta^2_{31} & \overline{q}_3\eta^2_{32} \\ -1 & 0 & -\overline{q}_4 & 0 \\ 0 & -1 & 0 & -\overline{q}_5 \end{array}\right] .$$

By the discussion after Theorem 1, the necessary and sufficient condition for geometric consistency is $\operatorname{rank} W_q = 5 - 4 + 2 = 3$, or, equivalently, that all 4×4-subdeterminants of W_q vanish. Consider for instance the subdeterminant obtained from the rows $1, 2, 3, 4$. Put $\zeta^1 = [\eta^1_{12}\ \eta^1_{22}\ \eta^1_{32}]^T$, $\zeta^2 = [\eta^2_{12}\ \eta^2_{22}\ \eta^2_{32}]^T$. It is readily verified that the subdeterminant condition can be written

$$\zeta^{1\,T} \Phi \zeta^2 = 0 \quad \text{with} \quad \Phi = \begin{bmatrix} 0 & B_1 & -B_2 \\ -B_1 & 0 & B_3 \\ B_2 & -B_3 & 0 \end{bmatrix} \operatorname{diag}(\overline{q}_1, \overline{q}_2, \overline{q}_3) ,$$

and

$$B_1 = \begin{vmatrix} \eta^1_{31} & \overline{q}_3\eta^2_{31} \\ \eta^1_{41} & \overline{q}_4\eta^2_{41} \end{vmatrix} , \quad B_2 = \begin{vmatrix} \eta^1_{21} & \overline{q}_2\eta^2_{21} \\ \eta^1_{41} & \overline{q}_4\eta^2_{41} \end{vmatrix} , \quad B_3 = \begin{vmatrix} \eta^1_{11} & \overline{q}_1\eta^2_{11} \\ \eta^1_{41} & \overline{q}_4\eta^2_{41} \end{vmatrix} .$$

This shows that $\zeta^{1\,T} \Phi \zeta^2 = 0$ is a necessary condition for the points ζ^1 and ζ^2 in the respective images to match. This is the classical epipolar constraint, cf. [2], and the matrix Φ is the fundamental matrix with respect to this particular choice of frames. The factorization of Φ was discovered in [6], in a slightly different setting, where it was called the *reduced fundamental matrix*. In an analogous way, trilinear and multilinear forms appear by taking subdeterminants of W_q when $m > 2$.

In the case of exact data, the statement that W_q has rank 3 is equivalent to the vanishing of a number of appropriately chosen subdeterminants, some of which can be interpreted as fundamental matrices. However, using such a finite family of algebraic conditions in the presence of noise, there is no longer any guarantee for the fulfillment of the rank condition. The same objection remains in the case of multilinear forms, and depicts a drawback of algorithms based on fundamental and multilinear forms. Another disadvantage is the coordinate dependency, which may require rules for coordinate normalization.

All these problems are avoided by working with the matrix W_q and the Chasles matrix, where, loosely speaking, simultaneous and uniform averaging is done over all conceivable constraints. One is lead to a viewpoint, where there is nothing special with the epipolar constraint compared to the other constraints that can be drawn from Theorem 1, except that contrary to most of the others, the epipolar constraint has a nice geometric interpretation.

3.4 Proximity measures

Intending to work with linear algebraic methods instead of polynomial equations, a quantitative tool for comparison of the correlation of linear subspaces is needed. In fact, the shape and depth theorem (single view) and the structure theorem (multiple views) both make assertments about the intersection of linear subspaces. In the single view case, the condition is that the intersection space of $\alpha s(\mathcal{X})$ and $s(\mathcal{Y})$ coincides with $\alpha s(\mathcal{X})$, and in the multiple view case, the condition is that the $n-3$-dimensional subspaces $s(\mathcal{Y}^1), \ldots, s(\mathcal{Y}^m)$ intersect in an $n-4$-dimensional subspace. This leads to the formulation of the

> *General problem:* Measure the rate of c-dimensional coincidence between linear subspaces $V_1, \ldots, V_m$ of $\mathbb{R}^n$.

To construct such a measure, let P_V denote the orthogonal projection matrix onto V. Then there is a chain of equivalences,

$$x \in V_1 \cap \ldots \cap V_m \iff \tfrac{1}{m}(P_{V_1}x + \ldots + P_{V_m}x) = x \iff$$
$$x \text{ eigenvector with eigenvalue 1 of } M = \tfrac{1}{m}(P_{V_1} + \ldots + P_{V_m}) \ .$$

It follows that $V_1, \ldots, V_m$ intersect in a c-dimensional subspace if and only if the eigenspace corresponding to the eigenvalue 1 of M has dimension c. Hence the matrix $I - M$ has rank deficiency c. A natural measure of this rank deficiency is the c:th smallest eigenvalue of $I - M$. An equivalent choice, more suitable for convergence studies of the algorithms below, is the following *proximity measure*:

$$\pi(V_1, \ldots, V_m) = (\sum_{k=1}^{c} \lambda_k^2)^{1/2} \text{ where } \lambda_1 \leq \ldots \leq \lambda_n \text{ are eigenvalues of } I - M \ .$$

In connection with single and multiple view geometry, as described by the shape and depth theorem and the structure theorem, $V = s(\mathcal{Y})$ for some $\mathcal{Y}$. Taking an S-matrix for $\mathcal{Y}$ with orthogonal columns, the projection matrix can be written SS^T. Using these theorems, a complication is the unknown depth parameters that appear. Violating slightly the orthogonality claim, in the case of single views below we work with the matrix

$$M = \frac{1}{2}(\Delta S_{\mathcal{X}} S_{\mathcal{X}}^T \Delta + S_{\mathcal{Y}} S_{\mathcal{Y}}^T) \quad \text{with} \quad \Delta = \operatorname{diag} \alpha \ ,$$

and in the case of multiple views, with the matrix

$$M = \frac{1}{m}(S_{\mathcal{Y}_1} S_{\mathcal{Y}_1}^T + Q_2 S_{\mathcal{Y}_2} S_{\mathcal{Y}_2}^T Q_2 + \ldots + Q_m S_{\mathcal{Y}_m} S_{\mathcal{Y}_m}^T Q_m) \ ,$$

with $Q_i = \operatorname{diag}(q^i), \ i = 1, \ldots, m$. An analogous construction can be done with depth spaces instead of shape spaces, cf. [8].

4 Algorithms for projective reconstruction

4.1 Complete data, no occlusions

By the discussion above, the problem is to determine kinetic depth vectors q so that the m-image analogue of (1) is fulfilled for some $\mathcal{X}_{||}$. This can be done by the following algorithm, introduced in [18]. A dual version, using depth instead of shape spaces, leads to factorization methods generalizing the one of [20], cf. [8]. The algorithm reads:

1. take $q^i = \mathbf{1}$ for $i = 1, \dots, m$,
2. compute an estimate of $\mathcal{X}$ by means of multiple view proximity,
3. knowing an estimate of $\mathcal{X}$, compute for each image i an estimate of the kinetic depth vector α^i by means of single view proximity, and form the corresponding kinetic depth vector q^i,
4. goto 2 or STOP, according to some criterion.

It can be shown that the sequence formed by the successively computed values of the proximity measure, $(\pi_k)_1^\infty$, decreases and convergences to a local minimum of π, considered as a function of $q^1, \dots, q^m$. Also the successively computed kinetic depth values and reconstruction estimates converge. In the case of exact data, and sufficiently many images and points, there is a unique minimum, corresponding to the true values of kinetic depth and the true object configuration $\mathcal{X}$. Empirically, the algorithm convergences very rapidly, in 10–20 iterations.

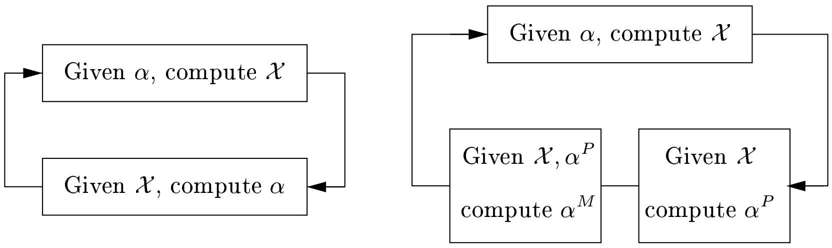

Fig. 1. Left loop: algorithm of Section 4.1. Right loop: algorithm of Secion 4.2.

The convergence can be proved by observing that the minimization problem hidden in the proximity measure can be formulated

$$\pi = \inf_{q^2,\dots,q^m} \inf_P ||M(q^2, \dots, q^m) - P||_{\text{Frob}}$$

where P runs through the orthogonal projection matrices of rank $n - 4$. This minimization problem can be studied by classical analytical methods.

4.2 Missing data, occlusions

The algorithm above can easily be modified to handle also the the situation of missing data, i.e. when not all points are visible in all images. In this case the second step of the algorithm in Section 4.1 is divided into two:

$3'$ knowing an estimate of $\mathcal{X}$, compute by means of single view proximity for each image i a depth vector $\alpha^{P,i}$ corresponding to points present in image i,
$3''$ compute for each image i the depth vector $\alpha^{M,i}$ corresponding to the missing points, using that the total depth vector $\alpha^i \in d(\mathcal{X})$.

The algorithms have been tested on images of a London scene, provided by Fraunhofer IGD within the CUMULI project. Six images have been used, taken at different locations on the river bank of the Thames. Forty points were manually detected in the images, where due to occlusions about 20 % of data was missing. The outcome is illustrated in Figure 2 and Figure 3, left diagram.

5 Using subsidiary information

5.1 Coplanarities

In man-made scenes, one often knows apriori that certain points are coplanar. The formalism of affine shape is well adapted to this situation. Suppose for instance that the first four points are coplanar. Then $s(\mathcal{X})$ contains an element where all components except the first four vanish (cf. *(i)* in Example A.1). Hence, for a given S-matrix, there is a column vector z such that

$$\begin{bmatrix} \times \\ \times \\ \times \\ \times \\ 0 \\ \vdots \\ 0 \end{bmatrix} = \underbrace{\begin{pmatrix} \times & \cdots & \times \\ \times & \cdots & \times \\ \times & \cdots & \times \\ \times & \cdots & \times \\ \times & \cdots & \times \\ \vdots & \ddots & \vdots \\ \times & \cdots & \times \end{pmatrix}}_{S_{\mathcal{X}}} z \ .$$

In case of non-exact data, this can't be expected to be fulfilled exactly. However, by a least square argument, the S-matrix can easily be adjusted to fulfill one or several coplanarity conditions. For instance, in the situation above, let ξ be the element in $s(\mathcal{X})$ that is closest to the linear space U consisting of vectors with vanishing components $5, \dots, n$, and let ξ' be the projection of ξ on U. Let $\xi^\perp$ be the orthogonal complement of ξ in $s(\mathcal{X})$. Then $\xi' \oplus \xi^\perp$ is the subspace of Σ_0 that is closest to $s(\mathcal{X})$ in proximity measure π. It is shape space of some configuration $\mathcal{X}'$, obeying the coplanarity constraint. In the same way, multiple coplanarity constraints can be handled.

This leads to an algorithm, illustrated by the left hand loop in Figure 4, yielding projective reconstruction of an object fulfilling a family of coplanarity constraints.

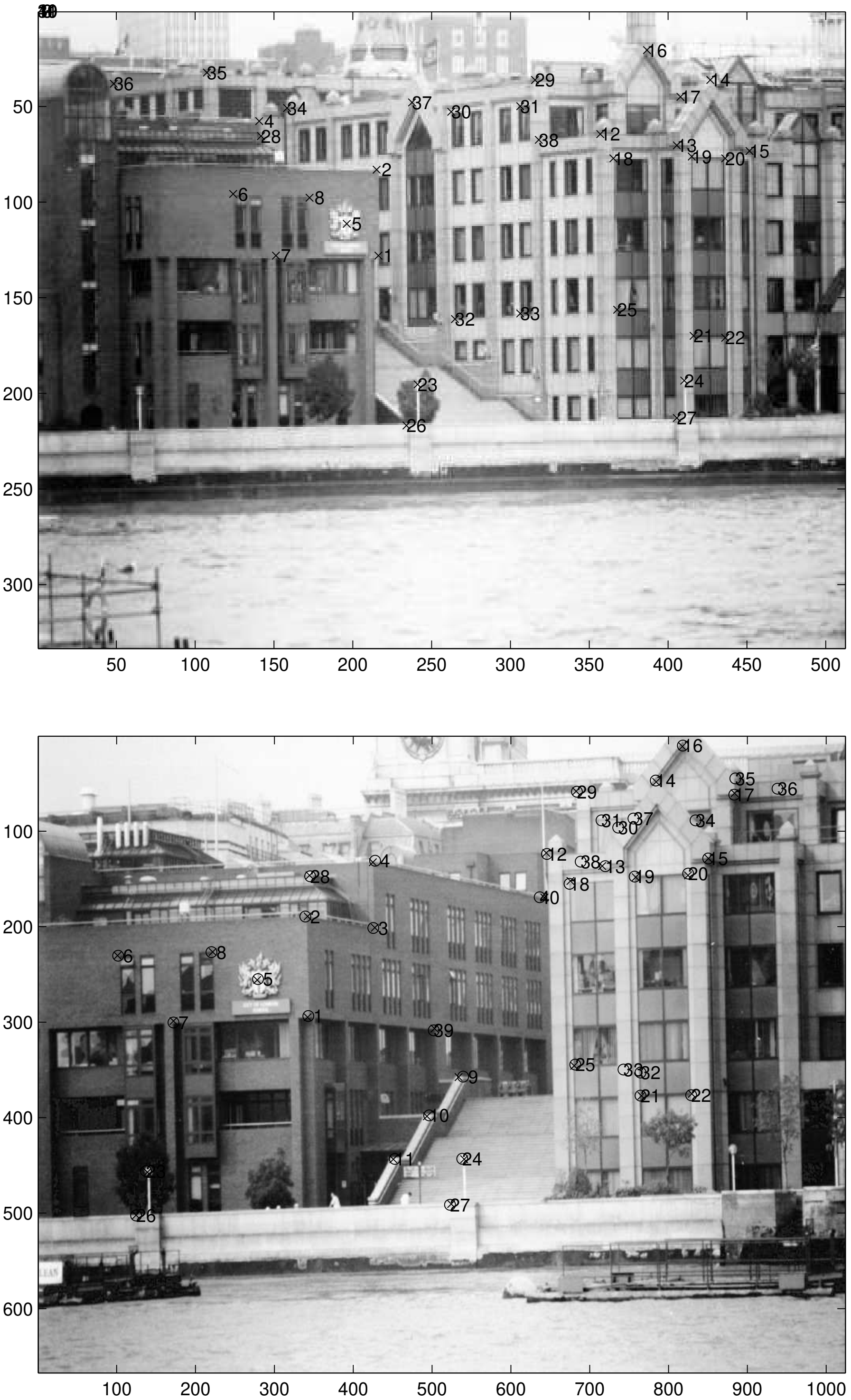

Fig. 2. Two of the six images used of a London scene. The symbols $\times$ denote the point configuration used, consisting of 40 points, with lots of occlusions. The circle symbols denote backprojection of reconstructed points, as if the scene had been transparent.

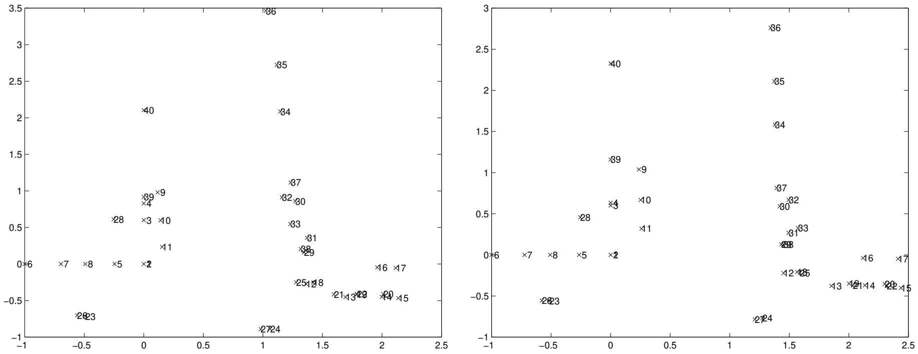

Fig. 3. Bird-eye perspectivities of the London scene, with placement of four points on the left building according to a city map. Left image: Projective reconstruction without using subsidiary information. Right image: Affine reconstruction, corrected for coplanarities given by the walls, and parallelities given by roofs, walls, windows and ground.

5.2 Parallelity

Often one knows not only that certain points A, B, C, D are coplanar, but also that some lines AB and CD connecting them are parallel. If sufficiently many such parallelities are known, the projective reconstruction can be strengthened to an affine one. In fact, it is easily seen that (cf. *(ii)* in Example A.1) $AB \parallel CD$ if and only if

$$\begin{bmatrix} a \\ -a \\ b \\ -b \\ 0 \\ \vdots \\ 0 \end{bmatrix} \in \beta s(\mathcal{X}_{\parallel}) \quad \text{for some} \quad a, b \ .$$

This leads to a linear system of equations in β, from which the depth in the first image can be determined, yielding an affine reconstruction of the object configuration $\mathcal{X}$. If Euclidean coordinates of four points are known, then Euclidean reconstruction of the whole configuration is achieved. An algorithm is described by the left hand loop in Figure 4. The performance on the London images is illustrated in Figure 3, right diagram.

5.3 Known focii locations

By means of Theorem 2 it is also possible to use the affine shape formalism to make Euclidean hand-eye calibration, even in the case of uncalibrated cameras, provided that the focal points are known. In fact, knowing the affine shape of the configuration formed by five or more focal points, from the Chasles matrix the

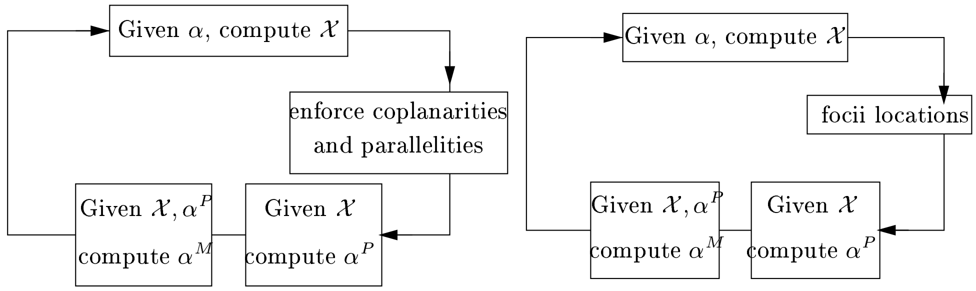

Fig. 4. Left loop: algorithm of Section 5.1 and 5.2. Right loop: algorithm of Secion 5.3.

depth in the first image can be computed, cf. [9], by iterative solving of linear systems of equations. This is done by a similar argument as the one behind the algorithm in Section 5.1, and yields an affine reconstruction. When knowing Euclidean coordinates for the focal points, from the Chasles matrix also the location of all objects points can be computed, yielding Euclidean reconstruction. An algorithm is described by the right hand loop in Figure 4. Figure 5 illustrates the typical performance of the algorithm on simulated data, in a situation where the focal points are densely distributed far away from the object. It is interesting to note that the impact of image noise mainly consists in a translation of the object along the ray of sight, while its shape is preserved to a large extent.

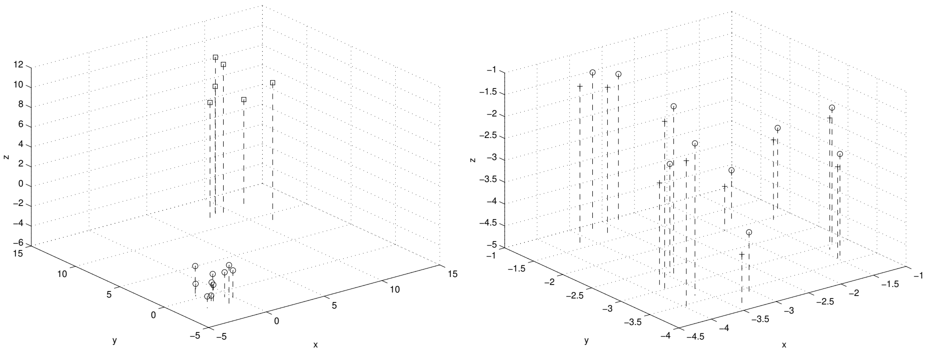

Fig. 5. Left diagram depicts focal points by squares and object points by circles. Right diagrams depicts true object points by circles and reconstructed points by crosses, in the presence of image noise.

References

1. Carlsson, S.: The Double Algebra: An Effective Tool for Computing Invariants in

Computer Vision. Applications of Invariance in Computer Vision, Springer Verlag, Lecture Notes in Computer Science **825** (1992) 145-164

2. Faugeras, O.D.: Three Dimensional Computer Vision, A Geometric Viewpoint. The MIT Press, Cambridge, Massachusetts, London (1993)
3. Faugeras, O.D. and Mourrain, B.: On the geometry and algebra of the point and line correspondences between n images, Proc. 5th ICCV, Cambridge, Massachusetts, IEEE Computer Society Press (1995) 951-956
4. Hartley, R. and Gupta, R. and Chang, T.: Stereo from uncalibrated cameras. Proc. IEEE Conf. on Computer vision and Pattern Recognition (1992) 761-764
5. Heyden, A.: Geometry and Algebra of Multiple Projective Transformations. PhD thesis, Dept of Mathematics, Lund (1995)
6. Heyden, A.: Reconstruction from Image Sequences by means of Relative Depths. Proc. 5th ICCV, Cambridge, Massachusetts, IEEE Computer Society Press (1995) 1058-1063
7. Heyden, A. and Åström, K.: A Canonical Framework for Sequences of Images. IEEE Workshop on Representation of Visual Scenes, IEEE Computer Society Press (1995)
8. Heyden, A.: Projective Structure and motion from image sequences using subspace methods. Proc. 10th Scandinavean Conference on Computer Vision (1997 963-968
9. Lindström, P.: Computation of geometric structure from an uncalibrated camera on a robot arm. ISRN LUTFD2/TFMA–98/7003–SE, Dept. of Mathematics, Lund (1998)
10. Shashua, A.: Trilinearity in Visual Recognition by Alignment. Computer Vision - ECCV94, Springer Verlag, Lecture Notes in Computer Vision **801** (1994) 479-484
11. Shashua, A.: Algebraic functions for recognition. IEEE Trans. Pattern Anal. Machine Intell. **17:8** (1995)
12. Sparr, G.: Projective invariants for affine shapes of point configurations. ESPRIT/DARPA Invariants Workshop, Reykjavik (1991)
13. Sparr, G: Depth computations from polyhedral images. Image and Vision Computing **10:10** (1992) 683-688
14. Sparr, G.: An algebraic/analytic method for reconstruction from image correspondences. In Theory & Applications of Image Analysis, Machine Perception Artificial Intelligence, World Scientific Publishing Co (Selected papers from SCIA'91) **2** (1992) 87-98
15. Sparr, A. and Sparr, G.: On a theorem of M. Chasles. Dept. of Mathematics, Lund, ISRN LUTFD2/TFMA–7001–SE (1993)
16. Sparr, G.: A common framework for kinetic depth, motion and reconstruction. Computer Vision – ECCV'94, Springer Verlag, Lecture Notes in Computer Science, Lecture Notes in Computer Science **801** (1994) 471-482
17. Sparr, G.: Structure and motion from kinetic depth. Dept. of Mathematics, Lund, ISRN LUTFD2/TFMA–95/7016–SE (Lecture notes from Sophus Lie symposium on Computer Vision, Nordfjordeid) (1995)
18. Sparr, G.: Simultaneous reconstruction of scene structure and camera locations from uncalibrated image sequences. Proc. ICPR'96 (1996) A328-333
19. Spetsakis, M. and Aloimonos, Y.: A unified theory of structure from motion. DARPA IU Workshop (1990) 271-283
20. Tomasi, C. and Kanade, T.: Shape and motion from image streams under orthography: A factorization method. Int. Journal of Computer Vision, **9:2** (1992) 137-154
21. Triggs, B.: Matching Constraints and the Joint Image. Proc. 5th ICCV, Cambridge, Massachusetts, IEEE Computer Society Press (1995) 338-343

Appendix

A Affine Shape and Depth

In this section, a brief recapitulation of the definitions and the basic properties of affine shape and depth spaces is given. For more details and proofs, see [12], [13], [14], [15], [16], [17], [18].

A.1 Subspace formulation

Let $\mathbf{A}^d$ denote an affine space of dimension d, where the cases $d = 2$ and $d = 3$ are of particular interest. In our approach, the primitive objects are not individual points of $\mathbf{A}^d$, but *point configurations*,

$$\mathcal{X} = (X^1, \dots, X^n), \quad \text{where} \quad X^k \in \mathbf{A}^d,\ k = 1, \dots, n \ .$$

By the *dimension* $\rho_{\mathcal{X}}$ of $\mathcal{X}$ is meant the dimension of the smallest affine subspace containing $\mathcal{X}$. Let the set of n-point configurations of dimension ρ be denoted $\mathcal{C}_{n,\rho}$. For instance, $\mathcal{C}_{n,2}$ consists of n-point configurations in $\mathbf{A}^3$ which are planar but not linear.

The main idea of the approach of affine shape is to peel off any dependency of the coordinatization of $\mathbf{A}^d$ on the parametrization of $\mathcal{C}_{n,\rho}$. To construct such a parametrization, consider two different coordinate representations x and $\bar{x}$ on $\mathbf{A}^d$. Here $\bar{x} = Bx + b$, where B is a non-singular $d \times d$-matrix, and b is a column matrix. For a given configuration $\mathcal{X}$, to the respective coordinate representations we associate matrices with the coordinate vectors as columns, but augmented with a row of ones,

$$X_a = \begin{bmatrix} x^1 & \dots & x^n \\ 1 & \dots & 1 \end{bmatrix} \ , \quad \overline{X}_a = \begin{bmatrix} \bar{x}^1 & \dots & \bar{x}^n \\ 1 & \dots & 1 \end{bmatrix} \ .$$

Then

$$\overline{X}_a = AX_a \quad \text{with} \quad A = \begin{bmatrix} B & b \\ 0 & 1 \end{bmatrix} \ . \tag{3}$$

In this way, the set of augmented coordinate matrices is partitioned into equivalence classes, each of which can be identified with one particular point configuration. The problem is to label these equivalence classes. This is done by means of the two consequences of (3):

$$\mathcal{N}(\overline{X}_a) = \mathcal{N}(X_a) \ , \quad \mathcal{R}_{\text{row}}(\overline{X}_a) = \mathcal{R}_{\text{row}}(X_a) \ ,$$

where $\mathcal{N}$ stands for 'nullspace' (column) and $\mathcal{R}_{\text{row}}$ for 'row space'. We have seen that these linear subspaces discriminate between point configurations. On the other hand, one readily verifies that if $\mathcal{N}(\overline{X}) = \mathcal{N}(X)$ or $\mathcal{R}_{\text{row}}(\overline{X}) = \mathcal{R}_{\text{row}}(X)$, then there exists an affine transformation A such that $\overline{X}_a = AX_a$. This shows that the linear subspaces $\mathcal{N}(X_a)$ and $\mathcal{R}_{\text{row}}(X_a)$ stand in a one-one correspondence with the set of point configurations. Since they are independent of the coordinatization of $\mathbf{A}^d$, the following definition makes sense.

Definition A.1. *Let X_a be an augmented coordinate matrix for $\mathcal{X} \in \mathcal{C}_{n,\rho}$ with respect to some coordinate system. Then*

the affine shape space *of $\mathcal{X}$, denoted $s(\mathcal{X})$, is defined by $s(\mathcal{X}) = \mathcal{N}(X_a)$,*
the affine depth space *of $\mathcal{X}$, denoted $d(\mathcal{X})$, is defined by $d(\mathcal{X}) = \mathcal{R}_{\text{row}}(X_a)$.*

Often we use abbreviated denominations, saying e.g. 'affine shape', 'shape space' or simply 'shape' instead of 'affine shape space', analogously for depth. The denomination 'affine depth' will be motivated in Remark 1 below. Affine shape also has an interpretation in terms of barycentric coordinates, cf. e.g. [13].

The discussion above is summarized in the following theorem, saying that affine shape and affine depth are complete affine invariants.

Theorem A.1. *Let $\mathcal{X}, \overline{\mathcal{X}} \in \mathcal{C}_{n,\rho}$. The following statements are equivalent:*

- *$\mathcal{X}$ and $\overline{\mathcal{X}}$ can be mapped onto each other by an affine transformation,*
- *$s(\overline{\mathcal{X}}) = s(\mathcal{X})$,*
- *$d(\overline{\mathcal{X}}) = d(\mathcal{X})$.*

To continue, some further notations are needed. Let

$$\Sigma_0 = \{\xi \in (\xi_1, \dots, \xi_n) \in \mathbb{R}^n \mid \sum_{k=1}^{n} \xi_k = 0\} ,$$

and let a multiplication on $\mathbf{R}^n$ be defined by

$$\alpha\xi = (\alpha_1\xi_1, \dots, \alpha_n\xi_n) \quad \text{if} \quad \alpha = (\alpha_1, \dots, \alpha_n),\ \xi = (\xi_1, \dots, \xi_n) .$$

In the same way, division ξ/α is defined by componentwise division, provided that $\alpha_i \neq 0,\ i = 1, \dots, n$. We use the notation $\overline{\alpha} = \mathbf{1}/\alpha$, where $\mathbf{1} = (1, \dots, 1) \in \mathbf{R}^n$. Finally, for the situation in Theorem A.1 we use the notation

$$\overline{\mathcal{X}} \stackrel{s}{=} \mathcal{X} \Longleftrightarrow \overline{\mathcal{X}} \text{ and } \mathcal{X} \text{ have equal shape} .$$

The following example is crucial for some common kinds of subsidiary information. It shows that shape spaces mirror a lot of qualitative information about point configurations.

Example A.1.

(i) Let $\mathcal{X} = (X^1, \dots, X^n)$. Then the sub-configuration $(X^{k_1}, X^{k_2}, X^{k_3}, X^{k_4})$ is planar if and only if

$$\xi_{k_1} X^{k_1} + \ldots + \xi_{k_4} X^{k_4} = 0 \quad \text{with} \quad \xi_{k_1} + \ldots + \xi_{k_4} = 0 .$$

Hence $s(\mathcal{X})$ contains an element $\xi = (\xi_1, \dots, \xi_n)$ where all components except $\xi_{k_1}, \dots, \xi_{k_4}$ vanish.

(ii) One readily verifies that in *(i)*, the vectors $\overline{X^{k_1} X^{k_2}}$ and $\overline{X^{k_3} X^{k_4}}$ are parallel if and only if $\xi_{k_1} = -\xi_{k_2}, \xi_{k_3} = -\xi_{k_4}$. In particular, $(X^{k_1}, X^{k_2}, X^{k_3}, X^{k_4})$ forms a parallelogram if and only if $\pm\xi_{k_1} = \pm\xi_{k_2} = \pm\xi_{k_3} = \pm\xi_{k_4}$, with two positive and two negative signs.

(iii) It can be shown that the vector $\overline{X^{k_1} X^{k_2}}$ is parallel to the plane spanned by the points $X^{k_3}, X^{k_4}, X^{k_5}$ if and only if $s(\mathcal{X})$ contains an element $\xi = (\xi_1, \dots, \xi_n)$, where all components except $\xi_{k_1}, \dots, \xi_{k_5}$ vanish and $\xi_{k_1} = -\xi_{k_2}$.

Theorem A.2. *Let $\mathcal{X} \in \mathcal{C}_{n,\rho}$. Then*

- *the affine shape space fulfills*
  - $\dim s(\mathcal{X}) = n - \rho - 1$,
  - $s(\mathcal{X}) \subset \Sigma_0$,
- *the affine depth space fulfills*
  - $\dim d(\mathcal{X}) = \rho + 1$,
  - $\mathbf{1} = (1, \dots, 1) \in d(\mathcal{X})$,
- *the affine shape and depth spaces are connected by*
  - $d(\mathcal{X}) s(\mathcal{X}) = 0$,
  - $s(\mathcal{X}) \oplus \ d(\mathcal{X}) = \mathbb{R}^n$.

The theorem says that the generic dimension of $s(\mathcal{X})$ is $n - 3$ for 2D-configurations, and $n - 4$ for 3D-configurations. In the same way, the generic dimension of $d(\mathcal{X})$ is 3 for 2D-configurations and 4 for 3D-configurations.

By the last item, the shape and depth spaces of an n-point configuration are orthogonal complements of each other in $\mathbb{R}^n$, $s(\mathcal{X})^0 = d(\mathcal{X})$, $d(\mathcal{X})^0 = s(\mathcal{X})$. Although this makes one of them seem superfluous, it is practical to use them in parallel since they embody different aspects of the geometry, with shape space directed on point configurations and depth spaces on transformations.

The following theorem generalizes Theorem A.1 to non-singular transformations, typically from 3D to 2D.

Theorem A.3. *Let $\mathcal{X} \in \mathcal{C}_{n,\rho}$ and $\overline{\mathcal{X}} \in \mathcal{C}_{n,\rho-1}$. The following statements are equivalent:*

- *$\mathcal{X}$ can be mapped onto $\overline{\mathcal{X}}$ by an affine transformation,*
- $s(\mathcal{X}) \subset s(\overline{\mathcal{X}})$,
- $d(\overline{\mathcal{X}}) \subset d(\mathcal{X})$.

A.2 Matrix formulation

To make numerical computations, matrix representations of the linear spaces $s(\mathcal{X})$ and $d(\mathcal{X})$ are needed.

Definition A.2. *Let $\mathcal{X} \in \mathcal{C}_{n,\rho}$. Then*

- *by an S-matrix of $\mathcal{X}$ is meant a matrix with column space $s(\mathcal{X})$,*
- *by a D-matrix of $\mathcal{X}$ is meant a matrix with row space $d(\mathcal{X})$.*

Note that X_a is a D-matrix of $\mathcal{X}$. The following example illustrates some typical computations with S-matrices.

Example A.2. Let

$$S = \begin{bmatrix} 2 & 1 & 4 \\ -2 & 1 & 0 \\ 0 & -2 & -4 \\ 2 & 0 & 2 \\ -1 & -1 & -2 \\ -1 & 1 & 0 \end{bmatrix} .$$

We claim that this is an S-matrix of some configuration $\mathcal{X} \in \mathcal{C}_{6,2}$. In view of Theorem A.2, first note that the matrix fulfills the necessary conditions of having rank 3 and vanishing column sums. The column space is unaffected by multiplication from the right by a non-singular matrix. In particular, using the inverse of the submatrix of S formed by the rows 3, 4 and 5, we obtain a new matrix with the same column space,

$$-\begin{bmatrix} 2 & 1 & 4 \\ -2 & 1 & 0 \\ 0 & -2 & -4 \\ 2 & 0 & 2 \\ -1 & -1 & -2 \\ -1 & 1 & 0 \end{bmatrix} \begin{bmatrix} 0 & -2 & -4 \\ 2 & 0 & 2 \\ -1 & -1 & -2 \end{bmatrix}^{-1} = \begin{bmatrix} \frac{1}{2} & -1 & 0 \\ \frac{1}{2} & 1 & 0 \\ -1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & -1 \\ 0 & 1 & 1 \end{bmatrix} .$$

The corresponding elements of the columns of the matrix on the right hand side provide the barycentric coordinate representations for the points X^3, X^4 and X^5, respectively, with respect to the affine frame X^1, X^2, X^6. Here in fact more can be said. Thus the first column says that,

$$X^3 = \frac{1}{2}X^1 + \frac{1}{2}X^2 ,$$

which means that X^1, X^2, X^3 are collinear, and that X^3 is the centroid of X^1 and X^2. The second column says that

$$-X^1 + X^2 - X^3 + X^4 = 0 ,$$

which means that X^1, X^2, X^3, X^4 are vertices of a parallelogram. Finally, the third column says that

$$-X^5 + X^6 = 0 ,$$

i.e. that the points X^5 and X^6 coincide.

So far, we have avoided 'points at infinity'. The following example illustrates how they can be treated within the framework of affine shape.

Example A.3. Let X^1, X^2, X^3 be three fixed points, and let X^4 be defined by

$$\overline{X^1X^4} = w\xi_2\overline{X^1X^2} + w\xi_3\overline{X^1X^3} . \tag{4}$$

Then $s(\mathcal{X})$ is a one-dimensional subspace of $\mathbb{R}^4$, generated by the vector $(1 - w\xi_2 - w\xi_3, w\xi_2, w\xi_3, -1)$. Letting $w \longrightarrow \infty$, by (4) we are led to interpret X^4 as the point at infinity in the direction $\xi_2\overline{X^1X^2} + \xi_3\overline{X^1X^3}$. Taking limits also of $s(\mathcal{X})$, where $\mathcal{X} = (X^1, \ldots, X^4)$, one finds that

$$s(\mathcal{X}) = \{(\xi_1, \xi_2, \xi_3, 0) \quad \text{with} \quad \xi_1 + \xi_2 + \xi_3 = 0\} .$$

A.3 Relation to Grassman manifolds

During the last few years, Grassman-Cayley and exteriour algebra has attracted some attention in computer vision, cf. e.g. [1]. The discussion below aims at explaining the place of affine shape and depth in this context.

By Theorem A.2, every point configurations $\mathcal{X}$ obeys the inclusion $s(\mathcal{X}) \subset \Sigma_0$. Conversely, every linear subspace U of Σ_0 is shape space for some point configuration. In fact, if $\dim U = \sigma$, in the same way as in Example A.2 it is seen that $U = s(\mathcal{X})$ for some configuration $\mathcal{X}$ of dimension $\rho_{\mathcal{X}} = n - \sigma - 1$. By Theorem A.3 we thus have a one-to-one correspondence between linear subspaces of Σ_0 and point configurations modulo affine transformations,

$$\mathcal{C}_{n,\rho}/\mathrm{aff} \cong G(\Sigma_0, n - \rho - 1) \ ,$$

where $G(\Sigma_0, d)$ denotes the Grassman manifold consisting of all d-dimensional linear subspaces of $\Sigma_0 \subset \mathbf{R}^n$.

The connection to Grassman algebra can be made more precise. For instance, let $\mathcal{X}$ be a planar 4-point configuration, with an augmented coordinate matrix X_a. By definition, the S-matrix, which in this case only has one column, is obtained by solving for the nullspace of X_a. Cramer's rule yields the components

$$\xi_{ijk} = \det \begin{bmatrix} x_i & x_j & x_k \\ 1 & 1 & 1 \end{bmatrix} \ . \tag{5}$$

These are recognized as Plücker coordinates of the subspace $s(\mathcal{X})$ in $\mathbb{R}^4$. The same holds true for bigger configurations. For instance, if $\mathcal{X}$ is a 5-point configuration, then all $\times$:es in the S-matrix

$$S_{\mathcal{X}} = \begin{bmatrix} \times & \times \\ \times & \times \\ \times & \times \\ \times & 0 \\ 0 & \times \end{bmatrix}$$

are Plücker coordinates for $s(\mathcal{X})$.

From this we learn that in uncalibrated camera geometry, it is only parameters of the form (5) that matter. In fact, regardless of the choice of coordinate system, image data organize themselves into such packages. An important feature of the approach by affine shape is that the array structure of the S-matrix adds further geometric information, compared to the Plücker coordinates alone.

B Projective transformations and spaces

B.1 Background

By a *perspective transformation* $P : \mathbf{A}^3 \longrightarrow \mathbf{A}^2$ with *focus* ϕ and image plane Π is meant a transformation such that every point $X \in \mathbf{A}^3$ is mapped to the point of intersection of Π and the line ϕX. Perspective transformations are used to model the *pinhole camera*. Whenever the focus is of interest, the notation P_ϕ will be used.

To deal with perspective transformations and their compositions, the *projective* transformations, to get a coherent theory one has to adjoin *points at infinity* to the ambient affine space. As described in Example A.3, such points can be interpreted as directions in $\mathbf{A}^3$. This gives a model for the *d-dimensional projective space* $\mathbf{P}^d$. If ϕ is a point at infinity, then P_ϕ is a *parallel projection* in the direction described by ϕ.

If $Y = P_\phi(X)$, then $\overline{\phi X} = \alpha\overline{\phi Y}$ for some α, called the *depth* of X with respect to Y. If ϕ is a point at infinity, the depth is by definition 1. If a configuration $\mathcal{X}$ is mapped onto a configuration $\mathcal{Y}$ by a perspective transformation, and the depth of X^k with respect to Y^k is α_k, $k = 1, \ldots, n$, then the vector $\alpha = (\alpha_1, \ldots, \alpha_n) \in \mathbf{R}^n$ is called the *depth vector* of $\mathcal{X}$ with respect to $\mathcal{Y}$.

B.2 Shape, depth and projective transformations

The following theorem gives a complete description of the single view geometry.

Theorem B.1. Shape and depth theorem.

(i) If $\mathcal{X}, \mathcal{Y} \in \mathcal{C}_{n,\rho}$, then the following statements are equivalent:
- *there exists a perspective transformation P, such that $P(\mathcal{X}) \stackrel{s}{=} \mathcal{Y}$ with depth vector α,*
- *$\alpha s(\mathcal{X}) = s(\mathcal{Y})$,*
- *$\alpha d(\mathcal{Y}) = d(\mathcal{X})$.*

(ii) If $\mathcal{X} \in \mathcal{C}_{n,\rho}$, $\mathcal{Y} \in \mathcal{C}_{n,\rho-1}$, then the following statements are equivalent:
- *there exists a perspective transformation P, such that $P(\mathcal{X}) \stackrel{s}{=} \mathcal{Y}$ with depth vector α,*
- *$\alpha s(\mathcal{X}) \subset s(\mathcal{Y})$,*
- *$\alpha d(\mathcal{Y}) \subset d(\mathcal{X})$.*

Remark 1. Now the terminology 'affine depth space' can be motivated, giving the answer to the question: Which depths can occur in conjunction with $\mathcal{X}$? From Theorem B.1 and the fact that every subspace of Σ_0 is a shape space, it follows that α is the depth of a perspective mapping acting on $\mathcal{X}$ if and only if $\alpha s(\mathcal{X}) \subset \Sigma_0$, i.e. $\alpha \in s(\mathcal{X})^0$. Since, by Theorem A.2, $s(\mathcal{X})^0 = d(\mathcal{X})$, the name 'depth space' for $d(\mathcal{X})$ is adequate.

B.3 Location of focal point

By Theorem B.1, there exists a projective transformation from 3D to 2D, $P_\phi : \mathcal{X} \to \mathcal{Y}$, with depth α, if and only if there holds a strict inclusion between linear subspaces, $\alpha s(\mathcal{X}) \subset s(\mathcal{Y})$. According to the following theorem, there is a one-to-one correspondence between ϕ and the set-difference $s(\mathcal{Y}) \setminus \alpha s(\mathcal{X})$, and it is possible to express ϕ in terms of $\mathcal{X}$ by an explicit formula. In formulating the theorem, a degenerate case called 'flat projection' has to be singled out, for details see [17]. If $\mathcal{X}$ is an n-point configuration and ϕ a point, then $(\mathcal{X}, \phi)$ denotes the $n+1$-point configuration formed by adjoining ϕ as an $n+1$:th point after the points of $\mathcal{X}$.

Theorem B.2. Focal point theorem. *Suppose that* $\alpha s(\mathcal{X}) \subset s(\mathcal{Y})$, *where* $\alpha \in \mathbb{R}^n$, $\mathcal{X} \in \mathcal{C}_{n,\rho}$, *and* $\mathcal{Y} \in \mathcal{C}_{n,\rho-1}$. *Then* $\mathcal{Y} = P_\phi(\mathcal{Y})$ *with a non-flat projection* P_ϕ *if and only if*

$$\phi = \sum_{k=1}^{n} \overline{\alpha}_k \eta_k X_k / \sum_{k=1}^{n} \overline{\alpha}_k \eta_k \ , \tag{6}$$

with $\eta \in s(\mathcal{Y}) \setminus \alpha s(\mathcal{X})$. *Analogously when* ϕ *is a point at infinity. In either case, the compound configuration* $(\mathcal{X}, \phi)$ *has an* S*-matrix*

$$S_{(\mathcal{X},\phi)} = \begin{bmatrix} \mathrm{diag}\,(\overline{\alpha})\mathrm{S}_{\mathcal{Y}} \\ -\overline{\alpha}^T S_{\mathcal{Y}} \end{bmatrix} \ . \tag{7}$$

B.4 Depth, shape and camera matrices

With x, y denoting object and image coordinates, respectively, the *camera matrix* P fulfills $P \begin{bmatrix} x \\ 1 \end{bmatrix} = \lambda \begin{bmatrix} y \\ 1 \end{bmatrix}$. For point configurations, it follows that

$$PX_a = Y_a \Lambda \ , \quad \text{with} \quad \Lambda = \mathrm{diag}\,(\lambda_1, \ldots, \lambda_n) \ .$$

The camera matrix of course depends on the coordinate systems used for the scene and the images.

From the equation $PX_a = Y_a\Lambda$ one reads out that each row of $Y_a\Lambda$ is a linear combination of the rows of X_a, with coefficients from P. It follows that $\lambda d(\mathcal{Y}) \subset d(\mathcal{X})$, where $\lambda = (\lambda_1, \ldots, \lambda_n)$. Conversely, if $\lambda d(\mathcal{Y}) \subset d(\mathcal{X})$ it can be shown that there exists a matrix P such that $PX_a = Y_a\Lambda$. This depicts the connection between camera matrices and depth and shape spaces, and that $\lambda = \alpha$ is the depth vector of the camera transformation.

Working with camera matrices, it is well known that the focus is obtained as the nullspace of the camera matrix. Theorem B.2 gives a novel characterization, having the advantage of providing an explicit formula for the focal point in terms of the object $\mathcal{X}$. To see the connection, take $\eta \in s(\mathcal{Y})$. From $PX_a\Lambda^{-1} = Y_a$, it follows that $PX_a\Lambda^{-1}\eta = Y_a\eta = 0$, which shows that $X_a\Lambda^{-1}\eta$ belongs to the nullspace of P, and thus is a focal point.