A General Theory of Single View Recognition – the Affine Case – with Applications to Indexing Image Databases for Content Based Retrieval *

by

P.F. Stiller, Texas A&M University

and

C.A. Asmuth and C.S. Wan, David Sarnoff Research Center

Abstract:

Using the theory of correspondences from algebraic geometry, we develop methods to relate 3-D objects to 2-D images and vice versa. In effect, we provide a very general framework for the use of geometric invariants in image recognition. At the most concrete level, our techniques yield a system of polynomial equations in variables which represent both the 3-D invariants of the features on an object and the 2-D invariants of features in an image. These equations will be satisfied if and only if the object can produce the image up to affine transformations of both the object and the image. The case of projective invariants will be dealt with in a forthcoming paper. The applications considered are to single view recognition and to indexing image databases for content based retrieval.

* Supported by the National Information Display Laboratory

§0. Introduction.

In this paper we develop a general mathematical framework which explains the relationship between certain collections of features on 3-D objects and the 2-D images they produce. Specifically, we are concerned with single-view recognition of 3-D arrangements of features such as points and lines. The general problem of single-view recognition has been characterized as the "holy grail" of computer vision (Weiss [9]) and was one of the inspirations for the use of geometric invariants in object recognition. Unfortunately, negative results of Burns, Weiss, and Riseman [1] show that there are no general-case view invariants. (For an alternate proof of this result see §6 below.) This makes the use of invariants for single-view recognition a more complex problem. While it is intuitively clear that a single 2-D view carries useful information about the original 3-D arrangement, the question of how to extract that information remains, to a large extent, an open one.

We shall show that much of this information can be characterized by a correspondence (in the sense of algebraic geometry). One of our goals is to explain this general idea and to give some explicit examples of the results that can be obtained. Questions of specific applications, implementation, and robustness are dealt with in a separate paper (Asmuth, Stiller, and Wan [7]). In addition, we will focus exclusively on the affine case, leaving the projective case for separate consideration (Asmuth, Stiller, and Wan [8]).

We adopt the point of view that a 3-D arrangement of features should be characterized by its 3-D affine or projective invariants, and that information from a single 2-D view should be expressed in the form of constraints on those invariants. Those 3-D objects whose features have invariants satisfying the constraints are then candidates for the object being viewed. Moreover, the constraints should be functions of the 2-D invariants of the features (points and/or lines) in the image. These constraints are shown to be relatively simple polynomial expressions in the combined set of variables: 3D invariants plus 2D invariants. A typical result, which the reader can turn to now, is Theorem 4.

The general method also works directly on the level of the 3-D and 2-D features without passing to invariants. However, the resulting equations are "invariant" in an appropriate sense and so reduce to those we present here.

In the affine case, our general formalism involves interpreting a fundamental set of

3-D affine invariants for an ordered set of n points in $\mathbf{R}^3$ (which are not all coplanar) as an $n-4$ dimensional linear subspace K^{n-4} of some $\mathbf{R}^{n-1}$. This yields a point in the $3n-12$ dimensional Grassmannian $Gr_{\mathbf{R}}(n-4, n-1)$. Similarly, a fundamental set of 2-D affine invariants for an ordered set of n points in $\mathbf{R}^2$ (not all collinear) is expressed as an $n-3$ dimensional linear subspace L^{n-3} of the same $\mathbf{R}^{n-1}$, which yields a point in the $2n-6$ dimensional Grassmannian $Gr_{\mathbf{R}}(n-3, n-1)$. The constraint which relates a 3-D feature set consisting of an n-tuple of points to an image consisting of an n-tuple of points is the incidence relation $K^{n-4} \subset L^{n-3} \subset \mathbf{R}^{n-1}$. In other words, those objects (feature sets) whose invariants in the form of the subspace K^{n-4} lie in the subspace L^{n-3} obtained from the image, are candidates for the object being viewed. We remark that the set of all objects (all K^{n-4}) capable of producing a given image (an L^{n-3}) defines a subvariety, denoted $\sum_{L^{n-3}}$, of dimension $n-4$ in $Gr_{\mathbf{R}}(n-4, n-1)$. $\sum_{L^{n-3}}$ is known as a Schubert cycle. Similarly, the set of all L^{n-3} containing a fixed K^{n-4} (i.e., the set of images produced by a fixed object) is a two-dimensional linear subvariety (a projective plane) in $Gr_{\mathbf{R}}(n-3, n-1)$, which we denote by $\sum_{K^{n-4}}$ (a related result appears in Jacobs [10]).

Finally this incidence correspondence between objects and images can be written down as a set of explicit polynomial constraints – either in terms of the $5n$ coordinates coming from the n 3-D points and the n 2-D points, or in terms of the $5n-20$ affine invariants ($3n-12$ invariants in 3-D and $2n-8$ invariants in 2-D). For the reader with less mathematical background who wants to see these ideas in a concrete setting, we recommend looking at some of the results in *§4. Examples* first.

As an application, we have used this scheme to index an image database for content based retrieval (see Asmuth, Stiller, and Wan [7]). In that application, we used feature sets consisting mainly of points and lines. Our approach also works well in situations where more than one view is available. Each view generates a separate constraint space, say L and L', and we find that $K = L \cap L'$. This means that K, and hence the 3-D affine invariants of our feature set, can be completely determined from two views, subject only to measurement error in the images. The methods are also flexible enough to be able to handle instances when particular points and/or lines are occluded.

§1. The Generalized Weak Perspective Projection Model.

Let A denote a matrix in $SO(3)$, i.e. a matrix with $\det A = 1$ and $A^T A = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$, so that A effects a rigid rotation of space. Also let $\xi = \begin{pmatrix} \xi_1 \\ \xi_2 \\ \xi_3 \end{pmatrix}$ denote a vector which we think of as providing a rigid translation of space.

The standard perspective object-to-image transformation $\pi_{A,\xi}$ takes the form:

$$\begin{pmatrix} u' \\ v' \end{pmatrix} = \pi'_{A,\xi}\left(\begin{pmatrix} x \\ y \\ z \end{pmatrix}\right) = \left(\frac{\lambda}{gx + hy + kz + \xi_3 + \lambda}\right) \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix} \left[A \begin{pmatrix} x \\ y \\ z \end{pmatrix} + \xi\right]$$

where λ is the focal length and the transformation

$$\begin{pmatrix} x \\ y \\ z \end{pmatrix} \longmapsto A \begin{pmatrix} x \\ y \\ z \end{pmatrix} + \xi \quad \text{where} \quad A = \begin{pmatrix} a & b & c \\ d & e & f \\ g & h & k \end{pmatrix}$$

is a rigid motion of space that reflects various views of the object. This is of course just the restriction to $\mathbf{R}^3$ of a special case of projection from a point P in projective space $\mathbf{P}^3$ onto a hyperplane $H \subset \mathbf{P}^3$ not containing P:

$$\pi_{H,P}\colon\ \mathbf{P}^3 - \{P\} \to \mathbf{P}^2.$$

(H is isomorphic to the projective plane $\mathbf{P}^2$.) For details, see Harris [3] pg. 34.

As an approximation to perspective transformations, one often uses the so-called weak perspective model (see, e.g., Clemens [2]):

$$\begin{pmatrix} u' \\ v' \end{pmatrix} = \pi_{A,\xi}\left(\begin{pmatrix} x \\ y \\ z \end{pmatrix}\right) = \left(\frac{\lambda}{w_0 + \lambda}\right) \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix} \left[A \begin{pmatrix} x \\ y \\ z \end{pmatrix} + \xi\right]$$

where w_0 is an "average" depth for the objects being viewed. Weak perspective is just orthogonal projection followed by scaling, once the view has been established.

In analyzing an image, we will be required to choose coordinates in the image plane without reference to the actual spatial coordinates. This means that the coordinates $\binom{u}{v}$ in the image will be related to $\binom{u'}{v'}$ by a transformation of the form

$$\begin{pmatrix} u \\ v \end{pmatrix} = s'\left(N \begin{pmatrix} u' \\ v' \end{pmatrix}\right) + \begin{pmatrix} \eta_1 \\ \eta_2 \end{pmatrix}$$

where s' is a positive scalar, N is a special orthogonal matrix, i.e. $NN^T = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$, $\det N = 1$, and $\eta = \binom{\eta_1}{\eta_2}$ is a translation.

Thus with the weak perspective model, we can assume that we are dealing with transformations of the form

$$\begin{pmatrix} u \\ v \end{pmatrix} = \pi_{N,\eta,A,\xi,s}\left(\begin{pmatrix} x \\ y \\ z \end{pmatrix}\right) = sN\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix}\left[A\begin{pmatrix} x \\ y \\ z \end{pmatrix} + \xi\right] + \eta$$

where $s \in \mathbf{R}$, $s > 0$ is a scaling factor, specifically $s = s'\left(\frac{\lambda}{\omega_0+\lambda}\right)$.

In this paper, we will allow a slightly more general set of transformations which we call *generalized weak perspective transformations*. Specifically, we allow $\pi_{A,\xi}$ to be followed by a general affine transformation of the image plane, $B' = \begin{pmatrix} a' & b' & \eta_1 \\ c' & d' & \eta_2 \\ 0 & 0 & 1 \end{pmatrix}$, with $a'd' - b'c' \neq 0$. In simpler terms, if an object produces an image, then from this point of view it can produce any affine transformation of that image. Thus we allow transformations

$$\pi_{A,\xi,B'}\colon\ \mathbf{R}^3 \longrightarrow \mathbf{R}^2$$

of the form

$$\begin{pmatrix} u \\ v \\ 1 \end{pmatrix} = \pi_{A,\xi,B'}\left(\begin{pmatrix} x \\ y \\ z \\ 1 \end{pmatrix}\right) = B'\begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}\begin{pmatrix} A & \vdots & \xi \\ \cdots & \cdots & \cdots \\ 000 & \vdots & 1 \end{pmatrix}\begin{pmatrix} x \\ y \\ z \\ 1 \end{pmatrix}.$$

Finally, suppose we apply an arbitrary affine transformation of 3-space prior to projecting:

$$\pi_{A,\xi,B'} \circ B$$

where $B = \begin{pmatrix} a & b & c & \delta_1 \\ d & e & f & \delta_2 \\ g & h & i & \delta_3 \\ 0 & 0 & 0 & 1 \end{pmatrix}$ with $\det\begin{pmatrix} a & b & c \\ d & e & f \\ g & h & i \end{pmatrix} \neq 0$. We claim that the resulting transformation from $\mathbf{R}^3$ to $\mathbf{R}^2$ is again of the form $\pi_{\widetilde{A},\widetilde{\xi},\widetilde{B}'}$ for suitable $\widetilde{A}$, $\widetilde{\xi}$, and $\widetilde{B}'$. In other words, 3-D objects that are affinely equivalent will produce the same set of possible images, and with no loss of generality, we can take our generalized weak perspective transformations to be of the form

$$\pi_{B',B} = B'\begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}B$$

where B' is an affine transformation of 2-space and B is an affine transformation of 3-space.

The impact of these observations and choices is that, in a certain sense, the best we can hope to do is to relate objects (feature sets) up to affine transformation to images up to affine transformation. In other words, we can at best hope to relate 3-D affine invariants of sets of say points and lines, to the 2-D affine invariants of the images they produce and vice versa.

§2. Affine Invariants for Sets of Points

In this section, we introduce a new type of affine invariant for a set of points. Unlike the familiar numerical invariants commonly used in image recognition, this invariant is a linear subspace of a particular vector space. It is in many respects the most natural invariant and is certainly more general and more robust than the standard numerical invariants; avoiding as it does, the need for any special general position assumptions. We also show that the numerical invariants can be completely recovered from the subspace.

Let $P_i = (x_i, y_i, z_i)$ for $i = 0, \ldots, n-1$, be an ordered set of n non-coplanar points in $\mathbf{R}^3$, and consider the $4 \times n$ matrix M given by

$$M = \begin{pmatrix} x_0 & x_1 & & x_{n-1} \\ y_0 & y_1 & & y_{n-1} \\ z_0 & z_1 & \ldots & z_{n-1} \\ 1 & 1 & & 1 \end{pmatrix}.$$

We associate to a 3-D object whose features set consists of the points $P_0, \ldots, P_{n-1}$ an $(n-4)$-dimensional linear subspace K^{n-4} of $\mathbf{R}^n$:

$$K^{n-4} = \{p = (p_0, \ldots, p_{n-1})^T \in \mathbf{R}^n \text{ such that } Mp = (0,0,0,0)^T\}.$$

The fact that K^{n-4} has dimension $n-4$ follows from the observation that at least one 4×4 minor of M has non-zero determinant because the points are not all coplanar. We will sometimes refer to K^{n-4} as the *key*.

Notice that if we apply an affine transformation T to our set of points, we obtain a new $4 \times n$ matrix

$$M' = \begin{pmatrix} a & b & c & \xi_1 \\ d & e & f & \xi_2 \\ g & h & k & \xi_3 \\ 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} x_0 & x_1 & & x_{n-1} \\ y_0 & y_1 & & y_{n-1} \\ z_0 & z_1 & \ldots & z_{n-1} \\ 1 & 1 & & 1 \end{pmatrix} = TM$$

but the subspace K^{n-4} <u>does not</u> change. Thus we can regard K^{n-4} as an "affine invariant". Moreover, since $K^{n-4} \subset H^{n-1} = \left\{p = (p_0, \ldots, p_{n-1})^T \in \mathbf{R}^n \text{ such that } \sum_{i=0}^{n-1} p_i = 0\right\}$, we can assign to our n-tuple of points the unique point determined by K^{n-4} in the Grassmannian, $Gr_{\mathbf{R}}(n-4, n-1)$, of $(n-4)$-planes in $(n-1)$-space. The space $Gr_{\mathbf{R}}(n-4, n-1)$ is a well understood manifold of dimension $3n-12$ (see Griffiths and Harris [5]).

We remark that any rational function (in the sense of algebraic geometry) on this Grassmannian provides a numerical affine invariant and conversely (as we shall see below). In other words, the field of rational functions on $Gr_{\mathbf{R}}(n-4, n-1)$ can be identified with all affine invariant expressions $\frac{p(x_0,\ldots,z_{n-1})}{q(x_0,\ldots,z_{n-1})}$ where p and q are polynomials in the variables x_i, y_i, z_i, $i = 0, \ldots, n-1$.

To explore this further, assume that P_0, P_1, P_2, P_3 are not coplanar. One can then show that K^{n-4} is spanned by the vectors $v_i = (p_0^{(i)}, \ldots, p_{n-1}^{(i)})^T$, $i = 4, \ldots n-1$, where

$$
\begin{aligned}
p_0^{(i)} &= -\det \begin{pmatrix} x_1 & x_2 & x_3 & x_i \\ y_1 & y_2 & y_3 & y_i \\ z_1 & z_2 & z_3 & z_i \\ 1 & 1 & 1 & 1 \end{pmatrix} \Big/ \det \begin{pmatrix} x_0 & x_1 & x_2 & x_3 \\ y_0 & y_1 & y_2 & y_3 \\ z_0 & z_1 & z_2 & z_3 \\ 1 & 1 & 1 & 1 \end{pmatrix} \\
p_1^{(i)} &= \det \begin{pmatrix} x_0 & x_2 & x_3 & x_i \\ y_0 & y_2 & y_3 & y_i \\ z_0 & z_2 & z_3 & z_i \\ 1 & 1 & 1 & 1 \end{pmatrix} \Big/ \det \begin{pmatrix} x_0 & x_1 & x_2 & x_3 \\ y_0 & y_1 & y_2 & y_3 \\ z_0 & z_1 & z_2 & z_3 \\ 1 & 1 & 1 & 1 \end{pmatrix} \\
p_2^{(i)} &= -\det \begin{pmatrix} x_0 & x_1 & x_3 & x_i \\ y_0 & y_1 & y_3 & y_i \\ z_0 & z_1 & z_3 & z_i \\ 1 & 1 & 1 & 1 \end{pmatrix} \Big/ \det \begin{pmatrix} x_0 & x_1 & x_2 & x_3 \\ y_0 & y_1 & y_2 & y_3 \\ z_0 & z_1 & z_2 & z_3 \\ 1 & 1 & 1 & 1 \end{pmatrix} \\
p_3^{(i)} &= \det \begin{pmatrix} x_0 & x_1 & x_2 & x_i \\ y_0 & y_1 & y_2 & y_i \\ z_0 & z_1 & z_2 & z_i \\ 1 & 1 & 1 & 1 \end{pmatrix} \Big/ \det \begin{pmatrix} x_0 & x_1 & x_2 & x_3 \\ y_0 & y_1 & y_2 & y_3 \\ z_0 & z_1 & z_2 & z_3 \\ 1 & 1 & 1 & 1 \end{pmatrix} \\
p_i^{(i)} &= -1 \\
p_j^{(i)} &= 0 \quad \text{for} \quad j = 4, \ldots, i-1, i+1, \ldots, n-1.
\end{aligned}
$$

We immediately recognize the $3n-12$ expressions $p_j^{(i)}$, $i = 4, \ldots, n-1$ and $j = 1, 2, 3$ as the fundamental set of 3-D affine invariants for our n-points obtained by moving P_0, P_1, P_2, P_3 to (0,0,0), (1,0,0), (0,1,0), and (0,0,1) respectively, via an affine transformation.

We can map $G_{\mathbf{R}}(n-4, n-1)$ into projective space $\mathbf{P}_{\mathbf{R}}^{\binom{n-1}{n-4}-1}$ via the usual Plücker embedding. This gives us a set of global homogeneous coordinates to use in computations. Specifically, given a point in $G_{\mathbf{R}}(n-4, n-1)$, we take any $n-4$ vectors which span that $(n-4)$-plane in $H^{n-1} = \mathbf{R}^{n-1}$ and use them to form an $(n-4) \times (n-1)$ matrix. (This of course requires a choice of basis for H^{n-1} so that it can be identified with $\mathbf{R}^{n-1}$.) The determinants of the $\binom{n-1}{n-4}$ minors of size $(n-4) \times (n-4)$ provide the map into projective space $\mathbf{P}_{\mathbf{R}}^{\binom{n-1}{n-4}-1}$. Using the vectors above, we see that all $3n-12$ of the fundamental invariants appear as determinants of $(n-4)\times(n-4)$ minors (possibly up to sign). Thus the usual invariants can be recovered from the embedding

$$Gr_{\mathbf{R}}(n-4,n)\hookrightarrow \mathbf{P}_{\mathbf{R}}^{\binom{n-1}{n-4}-1}$$

by taking ratios of homogeneous coordinates on $\mathbf{P}_{\mathbf{R}}^{\binom{n-1}{n-4}-1}$.

Example: Let $P_0=(1,1,2), P_1=(0,-2,1), P_2=(3,1,1), P_3=(-1,0,3)$ and $P_4=(2,-1,2)$. The first four points are not coplanar and can be moved to (0,0,0), (1,0,0), (0,1,0) and (0,0,1) respectively by a unique affine transformation. This transformation carries P_4 to the point $(p,q,r)=(-\frac{1}{3},\frac{10}{3},3)$. These three values are the fundamental affine invariants of this configuration of five points.

Our subspace invariant $K^1\subset H^4\subset\mathbf{R}^5$ is the line $\{(15t,t,-10t,-9t,3t),\ t\in\mathbf{R}\}$. This is easily verified by noting that

$$\begin{pmatrix}1&0&3&-1&2\\1&-2&1&0&-1\\2&1&1&3&2\\1&1&1&1&1\end{pmatrix}\begin{pmatrix}15\\1\\-10\\-9\\3\end{pmatrix}=\begin{pmatrix}0\\0\\0\\0\end{pmatrix}.$$

This gives a point $(15:1:-10:-9:3)\in\mathbf{P}^4$ contained in the hyperplane defined by the sum of the homogeneous coordinates equaling zero. (This hyperplane is our copy of $Gr_{\mathbf{R}}(1,4)$ which is a projective three-space $\mathbf{P}^3$.) This point equals $(1-p-q-r:p:q:r:-1)$ in $\mathbf{P}^4$, so that the ratios of the homogeneous coordinates are in fact values of certain invariants. □

Now, if we apply a weak-perspective transformation and select coordinates in the image plane, we will get points $Q_0=(u_0,v_0),\ldots,Q_{n-1}=(u_{n-1},v_{n-1})$ which are the images of $P_0,\ldots,P_{n-1}$ respectively.

As in the 3-D case, we can define an $(n-3)$-dimensional linear subspace $L^{n-3}\subset H^{n-1}\subset\mathbf{R}^n$ by:

$$L^{n-3}=\left\{(q_0,\ldots,q_{n-1})^T \text{ s.t. } \begin{pmatrix}u_0&u_1&&u_{n-1}\\v_0&v_1&\ldots&v_{n-1}\\1&1&&1\end{pmatrix}\begin{pmatrix}q_0\\\vdots\\q_{n-1}\end{pmatrix}=\begin{pmatrix}0\\0\\0\end{pmatrix}\right\}.$$

L^{n-3} will sometimes be referred to as the *lookup*. This subspace can be viewed as a point in the Grassmannian of $(n-3)$-planes in $(n-1)$-space, $Gr_{\mathbf{R}}(n-3,n-1)$, which is a manifold of dimension $2n-6$ (see Harris [3]). As in the 3-D case, L^{n-3} is spanned by vectors whose components include a fundamental set of affine invariants for our 2-D point set. In fact, specifying an L^{n-3} is essentially equivalent to specifying a fundamental set of $2n-6$ affine invariants. Note that L^{n-3} can be defined for any ordered set of n non-collinear points in the plane.

Theorem 1. $K^{n-4} \subset L^{n-3}$ if L^{n-3} is obtained from the images $Q_0, \ldots, Q_{n-1}$ of the points $P_0, \ldots, P_{n-1}$ used to construct K^{n-4}.

Proof: K^{n-4} is invariant under an affine transformation of our point set. Thus we can assume that our generalized weak-perspective transformation takes the form

$$\begin{pmatrix} x \\ y \\ z \\ 1 \end{pmatrix} \longmapsto B' \begin{pmatrix} x \\ y \\ 1 \end{pmatrix}$$

for some 2-D affine transformation B'. L^{n-3} is then just the kernel of the transformation from $\mathbf{R}^n$ to $\mathbf{R}^3$ given by

$$B' \begin{pmatrix} x_0 & x_1 & & x_{n-1} \\ y_0 & y_1 & \ldots & y_{n-1} \\ 1 & 1 & & 1 \end{pmatrix}. \tag{3}$$

However, an affine change of coordinates in the image plane in no way affects L^{n-3}, because

$$\begin{pmatrix} u_0 & & u_{n-1} \\ v_0 & \ldots & v_{n-1} \\ 1 & & 1 \end{pmatrix} = \begin{pmatrix} a & b & \xi_1 \\ d & e & \xi_2 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} x_0 & & x_{n-1} \\ y_0 & \ldots & y_{n-1} \\ 1 & & 1 \end{pmatrix}$$

has the same kernel as $\begin{pmatrix} x_0 & & x_{n-1} \\ y_0 & \ldots & y_{n-1} \\ 1 & & 1 \end{pmatrix}$. From this it is obvious that $K^{n-4} \subset L^{n-3}$. □

Example: The points P_0, P_1, P_2, P_3, P_4 in the example above are moved to (0,2,3), $(1,-1,4)$, (2,3,3), $\left(-\frac{5}{3}, \frac{1}{3}, \frac{11}{3}\right)$, $\left(\frac{4}{3}, \frac{4}{3}, \frac{14}{3}\right)$ under the rigid motion given by

$$\begin{pmatrix} x \\ y \\ z \end{pmatrix} \longrightarrow \begin{pmatrix} \frac{2}{3} & -\frac{1}{3} & -\frac{2}{3} \\ \frac{2}{3} & \frac{2}{3} & \frac{1}{3} \\ \frac{1}{3} & -\frac{2}{3} & \frac{2}{3} \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix} + \begin{pmatrix} 1 \\ 0 \\ 2 \end{pmatrix}.$$

We project into the x, y-plane and scale by $\frac{1}{12}$. (The average depth is $\frac{11}{3}$ and we take a focal length of $\lambda = 1/3$.) This gives the points $\left(0, \frac{1}{6}\right)$, $\left(\frac{1}{12}, -\frac{1}{12}\right)$, $\left(\frac{1}{6}, \frac{1}{4}\right)$, $\left(-\frac{5}{36}, \frac{1}{36}\right)$, $\left(\frac{1}{9}, \frac{1}{9}\right)$.

In practice of course, coordinates in the image plane will be selected without regard to any global coordinates in space. So suppose our coordinates in the image plane differ from the above coordinates by the transformation

$$\begin{pmatrix} u \\ v \end{pmatrix} = \begin{pmatrix} 3 & 1 \\ -1 & 3 \end{pmatrix} \begin{pmatrix} u' \\ v' \end{pmatrix} + \begin{pmatrix} -1 \\ 1 \end{pmatrix}.$$

In other words, the points as measured in the image plane are $Q_0 = \left(-\frac{5}{6}, \frac{3}{2}\right)$, $Q_1 = \left(-\frac{5}{6}, \frac{2}{3}\right)$, $Q_2 = \left(-\frac{1}{4}, \frac{19}{12}\right)$, $Q_3 = \left(-\frac{25}{18}, \frac{11}{9}\right)$, $Q_4 = \left(-\frac{5}{9}, \frac{11}{9}\right)$. Our lookup subspace L^2 is then defined by:

$$\begin{pmatrix} -\frac{5}{6} & -\frac{5}{6} & -\frac{1}{4} & -\frac{25}{18} & -\frac{5}{9} \\ \frac{3}{2} & \frac{2}{3} & \frac{19}{12} & \frac{11}{9} & \frac{11}{9} \\ 1 & 1 & 1 & 1 & 1 \end{pmatrix} \begin{pmatrix} p_0 \\ p_1 \\ p_2 \\ p_3 \\ p_4 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}.$$

Notice that $K^1 \subset L^2$ as

$$\begin{pmatrix} -\frac{5}{6} & -\frac{5}{6} & -\frac{1}{4} & -\frac{25}{18} & -\frac{5}{9} \\ \frac{3}{2} & \frac{2}{3} & \frac{19}{12} & \frac{11}{9} & \frac{11}{9} \\ 1 & 1 & 1 & 1 & 1 \end{pmatrix} \begin{pmatrix} 15 \\ 1 \\ -10 \\ -9 \\ 3 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}. \qquad \square$$

We point out that an even stronger result is true, namely:

Theorem 2. *Given $P_0, \ldots, P_{n-1}$ in $\mathbf{R}^3$ not all coplanar and $Q_0, \ldots, Q_{n-1}$ in $\mathbf{R}^2$ not all collinear, then $Q_0, \ldots, Q_{n-1}$ is the image of $P_0, \ldots, P_{n-1}$ under some generalized weak perspective transformation if and only if $K^{n-4} \subset L^{n-3}$ where K^{n-4} and L^{n-3} are constructed from $P_0, \ldots, P_{n-1}$ and $Q_0, \ldots, Q_{n-1}$ respectively.*

Thus the incidence correspondence $K^{n-4} \subset L^{n-3} \subset H^{n-1}$ is the test of whether or not an object can produce a particular image or vice versa. Moreover, as we shall see, this correspondence can be described in terms of some explicit polynomial equations.

Since we are treating our 3-D objects as distinguishable only up to affine transformation, the space of essentially different objects is the set of all possible $K^{n-4} \subset H^{n-1} \subset \mathbf{R}^n$, namely $Gr_{\mathbf{R}}(n-4, n-1)$, which, as we remarked, is a $3n-12$ dimensional manifold that we will denote by X^{3n-12}. We somewhat loosely refer to X^{3n-12} as the *space of objects*. A subspace K^{n-4} is thought of as giving a point, denoted $[K^{n-4}]$, in X^{3n-12}.

We also distinguish images only up to affine transformation. The space of essentially different images is the set of all possible $L^{n-3} \subset H^{n-1} \subset \mathbf{R}^n$ which is the $2n-6$ dimensional Grassmannian $Gr_{\mathbf{R}}(n-3, n-1)$. We denote this space by Y^{2n-6} and loosely refer to it as the *space of images*. A subspace L^{n-3} is thought of as giving a point, denoted $[L^{n-3}]$, in Y^{2n-6}.

¿From this perspective, a fixed image L^{n-3} defines a subset $\sum_{L^{n-3}}^{n-4}$ of X^{3n-12}, namely:

$$\sum\nolimits_{L^{n-3}}^{n-4} = \{[K^{n-4}] \in X^{3n-12} \quad \text{such that} \quad K^{n-4} \subset L^{n-3}\}$$

which is just the subset of object space, X^{3n-12}, consisting of those "objects" that could have produced the image L^{n-3}. $\sum_{L^{n-3}}^{n-4}$ is clearly an $(n-4)$-dimensional submanifold of X^{3n-12} isomorphic to projective $(n-4)$-space, $\mathbf{P}^{n-4}$. It is an example of something called a Schubert cycle in $Gr_{\mathbf{R}}(n-4, n-1)$. Notice that two images L^{n-3} and $\tilde{L}^{n-3}$ either intersect in a unique K^{n-4} or contain no K^{n-4}. In the first case, there is only one "object" capable of producing both images and that object is $K^{n-4} = L^{n-3} \cap \tilde{L}^{n-3}$. Also the Schubert cycles $\sum_{L^{n-3}}^{n-4}$ and $\sum_{\tilde{L}^{n-3}}^{n-4}$ intersect in one point in X^{3n-12}. In the second case, the two images could not have been produced by any one object and $\sum_{L^{n-3}}^{n-4} \cap \sum_{\tilde{L}^{n-3}}^{n-4} = \emptyset$.

Now let's fix an object $[K^{n-4}] \in X^{3n-12}$ and ask for the set of all images it can produce. This is a subset $\sum_{K^{n-4}}^{2} \subset Y^{2n-6}$ consisting of all $[L^{n-3}] \in Y^{2n-6}$ such that $K^{n-4} \subset L^{n-3}$. This is another Schubert cycle. It is easily seen to be a projective plane, i.e.

$$\sum\nolimits_{K^{n-4}}^{2} \cong \mathbf{P}^2,$$

because to produce an L^{n-3} containing a fixed K^{n-4}, we must give a line in the orthogonal complement of K^{n-4} in H^{n-1}, which is a line in an $\mathbf{R}^3$. This result is a generalization of a result due to Jacobs' (Jacobs [10]). If we embed Y^{2n-6} in projective space via the standard Plücker embedding, $\sum_{K^{n-4}}^{2}$ will be a linear subvariety of Y^{2n-6}. In appropriate affine coordinates, we will get a plane in $\mathbf{R}^{2n-6}$. Of course not every plane in $\mathbf{R}^{2n-6}$ occurs. As Jacobs shows, and as we shall show in section four, the planes that arise are of the form $\ell \times \ell' \subset \mathbf{R}^{n-3} \times \mathbf{R}^{n-3} = \mathbf{R}^{2n-6}$ where ℓ and ℓ' are parallel lines in $\mathbf{R}^{n-3}$. To specify a pair of parallel lines in $\mathbf{R}^{n-3}$ requires $3n-12$ parameters, $2n-8$ for the first line and $n-4$ for a second line parallel to it. This of course matches exactly with our count; we have precisely a $3n-12$ dimensional family of planes $\sum^2_{K^{n-4}} \subset Y^{2n-6}$ as $[K^{n-4}]$ ranges over X^{3n-12}, because, as is easy to show, distinct K^{n-4}'s produce distinct $\sum^2_{K^{n-4}}$'s.

§3. Correspondences

In this section, we develop the abstract notion of a correspondence to explain the "many to one" and "one to many" nature of the relationship between objects (feature sets) and images. This means that the relationship can be captured by a variety in the sense of algebraic geometry (denoted Z below). Moreover, we can derive polynomial equations which can be used to test for the object/image relationship. The reader should be forewarned that this discussion is rather abstract. However, a specific example is completely worked out in the next section.

Suppose we have two spaces X and Y (which we assume are smooth varieties in the sense of algebraic geometry). A correspondence is given by a suitable subvariety Z of the product variety $X \times Y$. The picture is

$$Z \subset X \times Y$$

$$\pi_X \qquad\qquad \pi_Y$$

$$X \qquad\qquad\qquad Y$$

where π_X and π_Y are the natural projections $\pi_X(x, y) = x$ and $\pi_Y(x, y) = y$.

To each point $x_0 \in X$ we can associate a subvariety $Y_{x_0} = \pi_Y(\pi_X^{-1}(\{x_0\}) \cap Z)$ of Y. (Here $\pi_X^{-1}(\{x_0\}) = \{(x_0, y), y \in Y\}$.) Similarly to each point $y_0 \in Y$ we can associate a subvariety $X_{y_0} = \pi_X(\pi_Y^{-1}(\{y_0\} \cap Z)$ of X.

This is precisely the situation we faced above in relating objects to images. The space X becomes the "space of objects" $X^{3n-12} = Gr_{\mathbf{R}}(n-4, n-1)$ which is a smooth projective variety. The space Y becomes the space of images $Y^{2n-6} = Gr_{\mathbf{R}}(n-3, n-1)$. The correspondence is given by $Z^{3n-10} \subset X^{3n-12} \times Y^{2n-6}$ where Z^{3n-10} is the variety of related object-image pairs given by the simple incidence relation $K^{n-4} \subset L^{n-3}$:

$$Z^{3n-10} = \{([K^{n-4}], [L^{n-3}]) \quad \text{such that} \quad K^{n-4} \subset L^{n-3}\}.$$

Z^{3n-10} is a well-known example of a flag manifold (see Harris [3] pg. 148). It can easily be shown that Z^{3n-10} has dimension $3n - 10$. The picture becomes:

$$Z^{3n-10} \subset X^{3n-12} \times Y^{2n-6}$$

$$\pi_X \qquad \pi_Y$$

$$X^{3n-12} \qquad Y^{2n-6}$$

Since $X \times Y$ has total dimension $5n-18$, we expect Z to be locally described by $2n-8$ equations. These will (after much translation) be polynomials in the $5n-18$ fundamental invariants ($3n-12$ for the 3-D set of n points and $2n-6$ for the 2-D set of n points). These polynomial expressions will be zero if and only if the n 2-D points can be an image of the n 3-D points under a generalized weak perspective transformation.

Notice that $\pi_X^{-1}(\{x_0\}) \cap Z$ has dimension 2 – we expect this because varieties of codimension $3n-12$ and $2n-8$ generally intersect in something of codimension $5n-20$. In that case, the intersection will have dimension 2 in $X^{3n-12} \times Y^{2n-6}$. The projection of this intersection into Y yields $Y^2_{x_0} = \pi_Y(\pi_X^{-1}(\{x_0\}) \cap Z)$ which is just the previously discussed Schubert cycle $\sum^2_{K^{n-4}} \cong \mathbf{P}^2$, where $x_0 = [K^{n-4}] \in X^{3n-12}$. Likewise $\pi_Y^{-1}(\{y_0\}) \cap Z$ has dimension $n-4$ (codimension $2n-6$ plus $2n-8$ equals $4n-14$ and $(5n-18)-(4n-14)$ yields dimension $n-4$). The projection into X yields $X^{n-4}_{y_0} = \pi_X(\pi_Y^{-1}(\{y_0\}) \cap Z)$ which is just the Schubert cycle $\sum^{n-4}_{L^{n-3}} \cong \mathbf{P}^{n-4}$ where $y_0 = [L^{n-3}] \in Y^{2n-6}$.

In the next section, we will explicitly calculate the polynomials describing Z in a simple case. Before doing that however, let's consider "correspondences" as a way to relate $3D$ to $2D$ information in other settings. We omit many details as these will appear in future work (Asmuth, Stiller, and Wan [8]).

Consider $n \geq 5$ points in suitably general position in $\mathbf{R}^3$ and perspective projections of them to $\mathbf{R}^2$. Here our space of objects X^{3n-15} will have dimension $3n-15$. Our space of images Y^{2n-8} will have dimension $2n-8$. Our correspondence will be given by Z^{3n-12} of dimension $3n-12$

$$Z^{3n-12} \subset X^{3n-15} \times Y^{2n-8}$$

$$\pi_X \qquad \pi_Y$$

$$X^{3n-15} \qquad Y^{2n-8}$$

A given object $x_0 \in X^{3n-15}$ yields a subspace of possible images $Y^3_{x_0} \subset Y^{2n-8}$ of dimension 3. Each image $y_0 \in Y^{2n-8}$ can come from any object in a subspace $X^{n-4}_{y_0} \subset X^{3n-15}$ of dimension $n-4$.

For example, when $n = 6$, the diagram above becomes:

$$Z^6 \subset X^3 \times Y^4$$

$$\pi_X \qquad \pi_Y$$

$$X^3 \qquad\qquad Y^4$$

and Z will be locally described by *one* equation in the 7 projective invariants – 3 for the six 3-D points and 4 for the six 2-D points.

A word of caution is required here. In order for X^{3n-15} and Y^{2n-8} to be reasonable spaces (say, smooth quasi-projective varieties) some rather complicated general position assumptions may be required (see the notion of *stability* in Mumford [6]). Again, we treat this case in a forthcoming paper (Asmuth, Stiller, and Wan [8]).

As another example, consider the case of $n \geq 3$ lines in $\mathbf{R}^3$ (in suitably general position) under the action of the affine group. The space of objects X^{4n-12} has dimension $4n-12$, the space of images Y^{2n-6} has dimension $2n-6$, and the correspondence will be given by Z^{4n-10} of dimension $4n-10$ in $X^{4n-12} \times Y^{2n-6}$:

$$Z^{4n-10} \subset X^{4n-12} \times Y^{2n-6}$$

$$\pi_X \qquad \pi_Y$$

$$X^{4n-12} \qquad\qquad Y^{2n-6}$$

Each $x_0 \in X$ gives rise to a two-dimensional subspace of possible images $Y^2_{x_0} \subset Y^{2n-6}$ and each image $y_0 \in Y$ can come from objects in a $(2n-4)$-dimensional subspace $X^{2n-4}_{y_0} \subset X^{4n-12}$.

Notice that when $n = 4$ we get no information, since the space of images Y^2 has dimension 2. The first interesting case is therefore $n = 5$, when we have the diagram

$$
\begin{array}{ccc}
 & Z^{10} \subset X^8 \times Y^4 & \\
 \pi_X & & \pi_Y \\
X^8 & & Y^4
\end{array}
$$

This case is also discussed in Asmuth, Stiller, and Wan [8]. We turn now to an explicit example.

§4. Examples

In this section we will consider several examples which illustrate our general results. After studying these, the reader should have no trouble manufacturing examples for other combinations of features.

We begin with the simple case of five points $P_i = (x_i, y_i, z_i)$, $i = 0, \ldots 4$ in Euclidean three-space $\mathbf{R}^3$. The only general position assumption we make is that the points are not all coplanar. Later, when we give explicit formulas in terms of the more common expressions for the affine invariants, we will assume that P_0, P_1, P_2, P_3 are not coplanar.

Our 3-D affine invariant is the one-dimensional subspace K^1 of $H^4 \subset \mathbf{R}^5$ defined by

$$K^1 = \left\{ (p_0, p_1, p_2, p_3, p_4)^T \in \mathbf{R}^5 \text{ s.t. } \begin{pmatrix} x_0 & x_1 & x_2 & x_3 & x_4 \\ y_0 & y_1 & y_2 & y_3 & y_4 \\ z_0 & z_1 & z_2 & z_3 & z_4 \\ 1 & 1 & 1 & 1 & 1 \end{pmatrix} \begin{pmatrix} p_0 \\ p_1 \\ p_2 \\ p_3 \\ p_4 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \end{pmatrix} \right\}$$

where H^4 is the hyperplane in $\mathbf{R}^5$ defined by $\sum_{i=0}^{4} p_i = 0$. Knowledge of K^1 is precisely equivalent to knowledge of a fundamental set of three affine invariants for our set of five points ($3n - 12$ equals 3 when $n = 5$). Specifically, if we assume P_0, P_1, P_2, P_3 are not coplanar, then K^1 is spanned by the vector $\begin{pmatrix} \tilde{p}_0 \\ \tilde{p}_1 \\ \tilde{p}_2 \\ \tilde{p}_3 \\ \tilde{p}_4 \end{pmatrix}$ where

$$\tilde{p}_0 = -\det \begin{pmatrix} x_1 & x_2 & x_3 & x_4 \\ y_1 & y_2 & y_3 & y_4 \\ z_1 & z_2 & z_3 & z_4 \\ 1 & 1 & 1 & 1 \end{pmatrix} \Big/ \det \begin{pmatrix} x_0 & x_1 & x_2 & x_3 \\ y_0 & y_1 & y_2 & y_3 \\ z_0 & z_1 & z_2 & z_3 \\ 1 & 1 & 1 & 1 \end{pmatrix}$$

$$\tilde{p}_1 = \det \begin{pmatrix} x_0 & x_2 & x_3 & x_4 \\ y_0 & y_2 & y_3 & y_4 \\ z_0 & z_2 & z_3 & z_4 \\ 1 & 1 & 1 & 1 \end{pmatrix} \Big/ \det \begin{pmatrix} x_0 & x_1 & x_2 & x_3 \\ y_0 & y_1 & y_2 & y_3 \\ z_0 & z_1 & z_2 & z_3 \\ 1 & 1 & 1 & 1 \end{pmatrix}$$

$$\tilde{p}_2 = -\det\begin{pmatrix} x_0 & x_1 & x_3 & x_4 \\ y_0 & y_1 & y_3 & y_4 \\ z_0 & z_1 & z_3 & z_4 \\ 1 & 1 & 1 & 1 \end{pmatrix} \Big/ \det\begin{pmatrix} x_0 & x_1 & x_2 & x_3 \\ y_0 & y_1 & y_2 & y_3 \\ z_0 & z_1 & z_2 & z_3 \\ 1 & 1 & 1 & 1 \end{pmatrix}$$

$$\tilde{p}_3 = \det\begin{pmatrix} x_0 & x_1 & x_2 & x_4 \\ y_0 & y_1 & y_2 & y_4 \\ z_0 & z_1 & z_2 & z_4 \\ 1 & 1 & 1 & 1 \end{pmatrix} \Big/ \det\begin{pmatrix} x_0 & x_1 & x_2 & x_3 \\ y_0 & y_1 & y_2 & y_3 \\ z_0 & z_1 & z_2 & z_3 \\ 1 & 1 & 1 & 1 \end{pmatrix}$$

$$\tilde{p}_4 = -1$$

Notice that the affine transformation that sends $P_0, P_1, P_2,$ and P_3 to (0,0,0), (1,0,0), (0,1,0), and (0,0,1) respectively is:

$$(x,y,z) \longmapsto \left(\frac{\det\begin{pmatrix} x_0 & x & x_2 & x_3 \\ y_0 & y & y_2 & y_3 \\ z_0 & z & z_2 & z_3 \\ 1 & 1 & 1 & 1 \end{pmatrix}}{\det\begin{pmatrix} x_0 & x_1 & x_2 & x_3 \\ y_0 & y_1 & y_2 & y_3 \\ z_0 & z_1 & z_2 & z_3 \\ 1 & 1 & 1 & 1 \end{pmatrix}}, \frac{\det\begin{pmatrix} x_0 & x_1 & x & x_3 \\ y_0 & y_1 & y & y_3 \\ z_0 & z_1 & z & z_3 \\ 1 & 1 & 1 & 1 \end{pmatrix}}{\det\begin{pmatrix} x_0 & x_1 & x_2 & x_3 \\ y_0 & y_1 & y_2 & y_3 \\ z_0 & z_1 & z_2 & z_3 \\ 1 & 1 & 1 & 1 \end{pmatrix}}, \frac{\det\begin{pmatrix} x_0 & x_1 & x_2 & x \\ y_0 & y_1 & y_2 & y \\ z_0 & z_1 & z_2 & z \\ 1 & 1 & 1 & 1 \end{pmatrix}}{\det\begin{pmatrix} x_0 & x_1 & x_2 & x_3 \\ y_0 & y_1 & y_2 & y_3 \\ z_0 & z_1 & z_2 & z_3 \\ 1 & 1 & 1 & 1 \end{pmatrix}} \right),$$

so that the coordinates of P_4 under this transformation are $(\tilde{p}_1, \tilde{p}_2, \tilde{p}_3)$. Also notice that $\sum_{i=0}^{4} \tilde{p}_i = 0$ so that $\tilde{p}_0 = -\tilde{p}_1 - \tilde{p}_2 - \tilde{p}_3 + 1$. Moreover $\tilde{p}_1, \tilde{p}_2, \tilde{p}_3$ form a fundamental set of affine invariants in the sense that any affine invariant of five points is a rational function of these three functions of $x_0, y_0, \ldots, z_4$.

Alternatively, assuming only that P_0, P_1, P_2, P_3, P_4 are not coplanar, we have that K^1 is spanned by $\begin{pmatrix} p'_0 \\ p'_1 \\ p'_2 \\ p'_3 \\ p'_4 \end{pmatrix}$ where

$$p_0' = -\det \begin{pmatrix} x_1 & x_2 & x_3 & x_4 \\ y_1 & y_2 & y_3 & y_4 \\ z_1 & z_2 & z_3 & z_4 \\ 1 & 1 & 1 & 1 \end{pmatrix}$$

$$p_1' = \det \begin{pmatrix} x_0 & x_2 & x_3 & x_4 \\ y_0 & y_2 & y_3 & y_4 \\ z_0 & z_2 & z_3 & z_4 \\ 1 & 1 & 1 & 1 \end{pmatrix}$$

$$p_2' = -\det \begin{pmatrix} x_0 & x_1 & x_3 & x_4 \\ y_0 & y_1 & y_3 & y_4 \\ z_0 & z_1 & z_3 & z_4 \\ 1 & 1 & 1 & 1 \end{pmatrix}$$

$$p_3' = \det \begin{pmatrix} x_0 & x_1 & x_2 & x_4 \\ y_0 & y_1 & y_2 & y_4 \\ z_0 & z_1 & z_2 & z_4 \\ 1 & 1 & 1 & 1 \end{pmatrix}$$

$$p_4' = -\det \begin{pmatrix} x_0 & x_1 & x_2 & x_3 \\ y_0 & y_1 & y_2 & y_3 \\ z_0 & z_1 & z_2 & z_3 \\ 1 & 1 & 1 & 1 \end{pmatrix}.$$

The set of all possible K^1's is parameterized by the set of all one-dimensional subspaces in H^4 which is a projective three-space $\mathbf{P}^3$. Thus our space of "3-D objects" X^3 is the three-dimensional manifold $\mathbf{P}^3$. We can take the homogeneous coordinates on X^3 to be $(p_1' : p_2' : p_3' : p_4')$. The field of rational functions on this $\mathbf{P}^3$ can be identified with the field of affine invariant rational functions in the variables x_i, y_i, z_i for $i = 0, \ldots, 4$.

Caution: This choice of coordinates on H^4 means that we have identified H^4 with $\mathbf{R}^4$ via the maps φ_1 and φ_2:

$$H^4 = \left\{ (p_0, \ldots, p_4)^T \in \mathbf{R}^5 \text{ s.t. } \sum_{i=0}^{4} p_i = 0 \right\} \xrightarrow{\varphi_1} \mathbf{R}^4$$

where $\varphi_1((p_0, \ldots, p_4)^T) = (p_1, p_2, p_3, p_4)^T$, and

$$\mathbf{R}^4 \xrightarrow{\varphi_2} H^4$$

where $\varphi_2((a, b, c, d)^T) = (-a - b - c - d, a, b, c, d)^T$. In other words, we have chosen the vectors $(-1, 1, 0, 0, 0)$, $(-1, 0, 1, 0, 0)$, $(-1, 0, 0, 1, 0)$ and $(-1, 0, 0, 0, 1)$ as a basis for $H^4 \subset \mathbf{R}^5$. It is important to note that these vectors are *not* orthonormal as vectors in $\mathbf{R}^5$.

In this example, an image will consist of five non-collinear points $Q_i = (u_j, v_j)$, $j = 0, \ldots, 4$ in the plane $\mathbf{R}^2$. Our 2-D affine invariant is the two-dimensional subspace L^2 of $H^4 \subset \mathbf{R}^5$ defined by

$$L^2 = \left\{ (q_0, q_1, q_2, q_3, q_4)^T \in \mathbf{R}^5 \text{ s.t. } \begin{pmatrix} u_0 & u_1 & u_2 & u_3 & u_4 \\ v_0 & v_1 & v_2 & v_3 & v_4 \\ 1 & 1 & 1 & 1 & 1 \end{pmatrix} \begin{pmatrix} q_0 \\ q_1 \\ q_2 \\ q_3 \\ q_4 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix} \right\}.$$

As in the three dimensional case, knowledge of L^2 is equivalent to knowledge of a fundamental set of 4 affine invariants for our five points in the plane ($2n - 6$ equals 4 when $n = 5$).

If one assumes that Q_0, Q_1, Q_2 are not collinear, then one can show that L^2 is spanned by the two independent vectors $(q'_0, q'_1, q'_2, q'_3, q'_4)^T$ and $(q''_0, q''_1, q''_2, q''_3, q''_4)^T$ where

$$q'_0 = \det \begin{pmatrix} u_1 & u_2 & u_3 \\ v_1 & v_2 & v_3 \\ 1 & 1 & 1 \end{pmatrix} \qquad q''_0 = \det \begin{pmatrix} u_1 & u_2 & u_4 \\ v_1 & v_2 & v_4 \\ 1 & 1 & 1 \end{pmatrix}$$

$$q'_1 = -\det \begin{pmatrix} u_0 & u_2 & u_3 \\ v_0 & v_2 & v_3 \\ 1 & 1 & 1 \end{pmatrix} \qquad q''_1 = -\det \begin{pmatrix} u_0 & u_2 & u_4 \\ v_0 & v_2 & v_4 \\ 1 & 1 & 1 \end{pmatrix}$$

$$q'_2 = \det \begin{pmatrix} u_0 & u_1 & u_3 \\ v_0 & v_1 & v_3 \\ 1 & 1 & 1 \end{pmatrix} \quad \text{and} \quad q''_2 = \det \begin{pmatrix} u_0 & u_1 & u_4 \\ v_0 & v_1 & v_4 \\ 1 & 1 & 1 \end{pmatrix}$$

$$q'_3 = -\det \begin{pmatrix} u_0 & u_1 & u_2 \\ v_0 & v_1 & v_2 \\ 1 & 1 & 1 \end{pmatrix} \qquad q''_3 = 0$$

$$q'_4 = 0 \qquad q''_4 = -\det \begin{pmatrix} u_0 & u_1 & u_2 \\ v_0 & v_1 & v_2 \\ 1 & 1 & 1 \end{pmatrix}$$

or the two independent vectors $(\tilde{q}_0, \tilde{q}_1, \tilde{q}_2, \tilde{q}_3, \tilde{q}_4)^T$ and $(\tilde{\tilde{q}}_0, \tilde{\tilde{q}}_1, \tilde{\tilde{q}}_2, \tilde{\tilde{q}}_3, \tilde{\tilde{q}}_4)^T$ where

$$\tilde{q}_0 = \det\begin{pmatrix} u_1 & u_2 & u_3 \\ v_1 & v_2 & v_3 \\ 1 & 1 & 1 \end{pmatrix} \Big/ \det\begin{pmatrix} u_0 & u_1 & u_2 \\ v_0 & v_1 & v_2 \\ 1 & 1 & 1 \end{pmatrix} \qquad \tilde{\tilde{q}}_0 = \det\begin{pmatrix} u_1 & u_2 & u_4 \\ v_1 & v_2 & v_4 \\ 1 & 1 & 1 \end{pmatrix} \Big/ \det\begin{pmatrix} u_0 & u_1 & u_2 \\ v_0 & v_1 & v_2 \\ 1 & 1 & 1 \end{pmatrix}$$

$$\tilde{q}_1 = -\det\begin{pmatrix} u_0 & u_2 & u_3 \\ v_0 & v_2 & v_3 \\ 1 & 1 & 1 \end{pmatrix} \Big/ \det\begin{pmatrix} u_0 & u_1 & u_2 \\ v_0 & v_1 & v_2 \\ 1 & 1 & 1 \end{pmatrix} \qquad \tilde{\tilde{q}}_1 = -\det\begin{pmatrix} u_0 & u_2 & u_4 \\ v_0 & v_2 & v_4 \\ 1 & 1 & 1 \end{pmatrix} \Big/ \det\begin{pmatrix} u_0 & u_1 & u_2 \\ v_0 & v_1 & v_2 \\ 1 & 1 & 1 \end{pmatrix}$$

$$\tilde{q}_2 = \det\begin{pmatrix} u_0 & u_1 & u_3 \\ v_0 & v_1 & v_3 \\ 1 & 1 & 1 \end{pmatrix} \Big/ \det\begin{pmatrix} u_0 & u_1 & u_2 \\ v_0 & v_1 & v_2 \\ 1 & 1 & 1 \end{pmatrix} \qquad \tilde{\tilde{q}}_2 = \det\begin{pmatrix} u_0 & u_1 & u_4 \\ v_0 & v_1 & v_4 \\ 1 & 1 & 1 \end{pmatrix} \Big/ \det\begin{pmatrix} u_0 & u_1 & u_2 \\ v_0 & v_1 & v_2 \\ 1 & 1 & 1 \end{pmatrix}$$

$$\tilde{q}_3 = -1 \qquad \tilde{\tilde{q}}_3 = 0$$

$$\tilde{q}_4 = 0 \qquad \tilde{\tilde{q}}_4 = -1.$$

Notice that the affine transformation that takes Q_0 to (0,0), Q_1 to (1,0), and Q_2 to (0,1) is:

$$(u,v) \longmapsto \left(\frac{\det\begin{pmatrix} u_0 & u & u_2 \\ v_0 & v & v_2 \\ 1 & 1 & 1 \end{pmatrix}}{\det\begin{pmatrix} u_0 & u_1 & u_2 \\ v_0 & v_1 & v_2 \\ 1 & 1 & 1 \end{pmatrix}}, \frac{\det\begin{pmatrix} u_0 & u_1 & u \\ v_0 & v_1 & v \\ 1 & 1 & 1 \end{pmatrix}}{\det\begin{pmatrix} u_0 & u_1 & u_2 \\ v_0 & v_1 & v_2 \\ 1 & 1 & 1 \end{pmatrix}} \right)$$

so that the coordinates of Q_3 and Q_4 under this affine transformation are $(\tilde{q}_1, \tilde{q}_2)$ and $(\tilde{\tilde{q}}_1, \tilde{\tilde{q}}_2)$. Of course, the values $\tilde{q}_1$, $\tilde{q}_2$, $\tilde{\tilde{q}}_1$ and $\tilde{\tilde{q}}_2$ form a fundamental set of affine invariants for sets of five points in $\mathbf{R}^2$ (with the restrictive general position assumption that the first three be non-collinear), and any affine invariant of five points in the plane is a rational function of these four.

Our space of images Y^4 is just the space of all two dimensional linear subspaces L^2 of the four dimensional space $H^4 \cong \mathbf{R}^4$. Thus Y^4 is just the Grassmannian $Gr(2,4)$ which is a well-known four-dimensional manifold. The rational functions $\tilde{q}_1$, $\tilde{q}_2$, $\tilde{\tilde{q}}_1$, $\tilde{\tilde{q}}_2$ provide coordinates on a Zariski open set of $Gr(2,4)$. We can identify the function field of $Gr(2,4)$ with the field of all affine invariant rational functions in the variables $u_0, v_0, \ldots, u_4, v_4$.

In §2 above, we showed that an object, in this case described by a feature set consisting of five non-coplanar points, can produce a particular image, consisting of five non-collinear points, up to affine transformations of both the object and the image, if and only if the invariant one-dimensional subspace $K^1 \subset H^4 \subset \mathbf{R}^5$ associated to the object is contained in the invariant two-dimensional subspace $L^2 \subset H^4 \subset \mathbf{R}^5$ associated to the image. In other words $K^1 \subset L^2$ is a necessary and sufficient condition for a particular ordered set of five points in 3-D and a particular ordered set of five points in 2-D to be related in the sense that the 2-D points are the images of the 3-D points under a generalized weak perspective transformation.

For example, take $P_0 = (1,1,2)$, $P_1 = (0,-2,1)$, $P_2 = (3,1,1)$, $P_3 = (-1,0,3)$, $P_4 = (2,-1,2)$, $Q_0 = (-\frac{5}{6}, \frac{3}{2})$, $Q_1 = \big(-\frac{5}{6}, \frac{2}{3}\big)$, $Q_2 = (-\frac{1}{4}, \frac{19}{12})$, $Q_3 = \big(-\frac{25}{18}, \frac{11}{9}\big)$ and $Q_4 = (-\frac{5}{9}, \frac{11}{9})$. Then using the basis $(-1,1,0,0,0)$, $(-1,0,1,0,0)$, $(-1,0,0,1,0)$ and $(-1,0,0,0,1)$ for H^4, we have that K^1 is spanned by

$$\begin{pmatrix} \widetilde{p}_1 \\ \widetilde{p}_2 \\ \widetilde{p}_3 \\ \widetilde{p}_4 \end{pmatrix} = \begin{pmatrix} -1/3 \\ 10/3 \\ 3 \\ -1 \end{pmatrix}.$$

Note that as a five vector this is:

$$\widetilde{p}_1 \begin{pmatrix} -1 \\ 1 \\ 0 \\ 0 \\ 0 \end{pmatrix} + \widetilde{p}_2 \begin{pmatrix} -1 \\ 0 \\ 1 \\ 0 \\ 0 \end{pmatrix} + \widetilde{p}_3 \begin{pmatrix} -1 \\ 0 \\ 0 \\ 1 \\ 0 \end{pmatrix} + \widetilde{p}_4 \begin{pmatrix} -1 \\ 0 \\ 0 \\ 0 \\ 1 \end{pmatrix} = \begin{pmatrix} -5 \\ -1/3 \\ 10/3 \\ 3 \\ -1 \end{pmatrix}$$

so that $K^1 \subset \mathbf{R}^5$ is spanned by

$$\begin{pmatrix} 15 \\ 1 \\ -10 \\ -9 \\ 3 \end{pmatrix}$$

as indicated in §2 above. The point here is that $\widetilde{p}_1 = -1/3$, $\widetilde{p}_2 = 10/3$, and $\widetilde{p}_3 = 3$ are the standard affine invariants that one obtains by moving $P_0, \ldots, P_3$ to standard position and taking the coordinates of the point that P_4 moves to. Similarly L^2 is spanned by

$$\begin{pmatrix} \widetilde{q}_1 \\ \widetilde{q}_2 \\ \widetilde{q}_3 \\ \widetilde{q}_4 \end{pmatrix} = \begin{pmatrix} 5/21 \\ -20/21 \\ -1 \\ 0 \end{pmatrix} \quad \text{and} \quad \begin{pmatrix} \widetilde{\widetilde{q}}_1 \\ \widetilde{\widetilde{q}}_2 \\ \widetilde{\widetilde{q}}_3 \\ \widetilde{\widetilde{q}}_4 \end{pmatrix} = \begin{pmatrix} 8/21 \\ 10/21 \\ 0 \\ -1 \end{pmatrix}.$$

The values $(\widetilde{q}_1, \widetilde{q}_2) = (5/21, -20/21)$ and $(\widetilde{\widetilde{q}}_1, \widetilde{\widetilde{q}}_2) = (8/21, 10/21)$ are the standard affine invariants of $Q_0, \ldots, Q_4 \in \mathbf{R}^2$. They are the coordinates of Q_3 and Q_4 respectively, under the unique affine transformation that sends Q_0, Q_1, Q_2 to standard position, i.e. Q_0 to (0,0), Q_1 to (1,0), and Q_2 to (0,1).

Since our set of 2-D points is in fact an image of our set of 3-D points (under a generalized weak perspective transformation), we should have $K^1 \subset L^2$. This is indeed the case:

$$\begin{pmatrix} -1/3 \\ 10/3 \\ 3 \\ -1 \end{pmatrix} = -3 \begin{pmatrix} 5/21 \\ -20/21 \\ -1 \\ 0 \end{pmatrix} + \begin{pmatrix} 8/21 \\ 10/21 \\ 0 \\ -1 \end{pmatrix}.$$

The reader should also note that the expressions

$$(1) \qquad \widetilde{p}_1 + \widetilde{p}_3\widetilde{q}_1 - \widetilde{q}_3$$

and

$$(2) \qquad \widetilde{p}_2 + \widetilde{p}_3\widetilde{q}_2 - \widetilde{q}_4$$

are both zero where for simplicity we have changed notation replacing $(\widetilde{q}_1, \widetilde{q}_2, \widetilde{\widetilde{q}}_1, \widetilde{\widetilde{q}}_2)$ with $(\widetilde{q}_1, \widetilde{q}_2, \widetilde{q}_3, \widetilde{q}_4)$. These polynomials in the seven invariants $(\widetilde{p}_1, \widetilde{p}_2, \widetilde{p}_3)$ and $(\widetilde{q}_1, \widetilde{q}_2, \widetilde{q}_3, \widetilde{q}_4)$ give, as we shall see, the fundamental relationship between objects and images in the case of five points. An ordered set of five 3-D points with the first four not coplanar and an ordered set of five 2-D points with the first three not collinear are related by a generalized weak perspective transformation if and only if equations (1) and (2) are zero.

Two things to notice about these equations when they are set equal to zero. The first is that they are linear in each of the two sets of variables $(\widetilde{p}_1, \widetilde{p}_2, \widetilde{p}_3)$ and $(\widetilde{q}_1, \widetilde{q}_2, \widetilde{q}_3, \widetilde{q}_4)$. Fixing $(\widetilde{p}_1, \widetilde{p}_2, \widetilde{p}_3)$ yields a plane σ in $\mathbf{R}^4$ which is the product of two parallel lines ℓ_1 and ℓ_2 in $\mathbf{R}^2$, i.e.

$$\sigma = \ell_1 \times \ell_2 \subset \mathbf{R}^2 \times \mathbf{R}^2$$

where the coordinates on the first $\mathbf{R}^2$ are $(\widetilde{q}_1, \widetilde{q}_3)$ and the coordinates on the second $\mathbf{R}^2$ are $(\widetilde{q}_2, \widetilde{q}_4)$. The plane σ describes all the images that our fixed 3-D object can produce. This is just Jacobs' result in [2]. On the other hand, fixing the image, i.e. $\widetilde{q}_1, \widetilde{q}_2, \widetilde{q}_3, \widetilde{q}_4$, yields the equations of a line in $\mathbf{R}^3$. This is the set of all objects (five point feature sets up to 3-D affine transformation) that can produce our fixed image. The equations (1) and

(2) describe our correspondence Z^5 in an open subset, which can be identified with $\mathbf{R}^7$, of $X^3 \times Y^4$. (See §2 and §3 for an explanation of the notation.)

The second thing to notice is that the variables $\widetilde{p}_1, \widetilde{p}_2, \widetilde{p}_3$ are rational expressions in the coordinates of the 3-D points $P_i = (x_i, y_i, z_i)$ and that the variables $\widetilde{q}_1, \widetilde{q}_2, \widetilde{q}_3, \widetilde{q}_4$ are rational expressions in the coordinates of the 2-D points $Q_j = (u_j, v_j)$. After clearing denominators expressions (1) and (2) become bihomogeneous polynomials of bidegree 3,2 in the two sets of variables $\{x_i, y_i, z_i, i = 0, \ldots, 4\}$ and $\{u_j, v_j, j = 0, \ldots, 4\}$.

As we have shown, the condition that a particular "object" (feature set consisting of five points) with invariant subspace $K^1 \subset H^4 \subset \mathbf{R}^5$ can produce a particular image with invariant subspace $L^2 \subset H^4 \subset \mathbf{R}^5$ is given by $K^1 \subset L^2$. In order to write down the general equations for this relationship (assuming only that the five points in space are non-coplanar and the five points in the plane are non-collinear), we will need global coordinates on the space of objects $X^3 = \mathbf{P}^3$ and on the space of images $Y^4 = Gr(2,4)$.

For $X^3 = \mathbf{P}^3$ we use the homogeneous coordinates (p_1, p_2, p_3, p_4) where

$$p_1 = \det \begin{pmatrix} x_0 & x_2 & x_3 & x_4 \\ y_0 & y_2 & y_3 & y_4 \\ z_0 & z_2 & z_3 & z_4 \\ 1 & 1 & 1 & 1 \end{pmatrix}$$

$$p_2 = -\det \begin{pmatrix} x_0 & x_1 & x_3 & x_4 \\ y_0 & y_1 & y_3 & y_4 \\ z_0 & z_1 & z_3 & z_4 \\ 1 & 1 & 1 & 1 \end{pmatrix}$$

$$p_3 = \det \begin{pmatrix} x_0 & x_1 & x_2 & x_4 \\ y_0 & y_1 & y_2 & y_4 \\ z_0 & z_1 & z_2 & z_4 \\ 1 & 1 & 1 & 1 \end{pmatrix}$$

$$p_4 = -\det \begin{pmatrix} x_0 & x_1 & x_2 & x_3 \\ y_0 & y_1 & y_2 & y_3 \\ z_0 & z_1 & z_2 & z_3 \\ 1 & 1 & 1 & 1 \end{pmatrix}$$

are cubic polynomials in the coordinates of the five-tuple of 3-D points $P_i = (x_i, y_i, z_i)$. Keep in mind that $p_0 = -p_1 - p_2 - p_3 - p_4$ and that this choice of coordinates amounts to identifying H^4 with $\mathbf{R}^4$ via the specific choice of basis described earlier.

For $Y^4 = Gr(2,4)$ we get global coordinates by embedding the Grassmannian into $\mathbf{P}^5$ via the well-known Plücker embedding (see Appendix 1 below). Specifically, if $(M_{12} :$

$M_{13} : M_{14} : M_{23} : M_{24} : M_{34})$ are homogeneous coordinates on $\mathbf{P}^5$ and L^2 is a plane in $H^4 \cong \mathbf{R}^4$ spanned by $(a_1, a_2, a_3, a_4)^T$ and $(b_1, b_2, b_3, b_4)^T$ then

$$M_{ij} = \det \begin{pmatrix} a_i & a_j \\ b_i & b_j \end{pmatrix}.$$

The image of $Y^4 = Gr(2,4)$ in $\mathbf{P}^5$ is the quadric hypersurface cut out by

$$M_{12}M_{34} - M_{13}M_{24} + M_{14}M_{23} = 0. \tag{3}$$

Our object image correspondence Z^5, where

$$Z^5 \subset X^3 \times Y^4 = \mathbf{P}^3 \times Gr(2,4) \subset \mathbf{P}^3 \times \mathbf{P}^5,$$

will be described by a system of bihomogeneous polynomials in the two sets of variables $\{p_1, p_2, p_3, p_4\}$ and $\{M_{12}, M_{13}, M_{14}, M_{23}, M_{24}, M_{34}\}$. One equation will of course be (3), which is automatically satisfied by any image. Note that these coordinates are either 3-D or 2-D affine invariants up to scale, i.e. their ratio's are invariant. This is obvious in the case of the p_i, but less clear for the M_{ij}. To see this, we must compute the M_{ij} for a given image L^2. The problems here are that, first, we make no general position assumption except that our five image points are not all collinear, and that second, L^2 is described as the kernel of a linear map from $\mathbf{R}^5$ to $\mathbf{R}^3$.

Specifically, if $Q_j = (u_j, v_j)$, $j = 0, \ldots, 4$ are the five points in our image, then $(e, a, b, c, d)^T \in L^2$ if and only if

$$\begin{pmatrix} u_0 & u_1 & u_2 & u_3 & u_4 \\ v_0 & v_1 & v_2 & v_3 & v_4 \\ 1 & 1 & 1 & 1 & 1 \end{pmatrix} \begin{pmatrix} e \\ a \\ b \\ c \\ d \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}.$$

This is equivalent to

$$\begin{pmatrix} u_1 - u_0 & u_2 - u_0 & u_3 - u_0 & u_4 - u_0 \\ v_1 - v_0 & v_2 - v_0 & v_3 - v_0 & v_4 - v_0 \end{pmatrix} \begin{pmatrix} a \\ b \\ c \\ d \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

and $e = -a - b - c - d$.

The rows of this 2×4 matrix can be thought of as spanning a plane σ in $\mathbf{R}^4$. The invariant subspace L^2 associated with the image is $\sigma^\perp$. Notice that the Plücker coordinates of σ are $(N_{12}, N_{13}, N_{14}, N_{23}, N_{24}, N_{34})$ where $N_{ij} = \det \begin{pmatrix} u_0 & u_i & u_j \\ v_0 & v_i & v_j \\ 1 & 1 & 1 \end{pmatrix}$.

Using the result in Appendix 1, we find that the Plücker coordinates of $L^2 = \sigma^\perp$ are:

$$(M_{12}, M_{13}, M_{14}, M_{23}, M_{24}, M_{34}) = (N_{34}, -N_{24}, N_{23}, N_{14}, -N_{13}, N_{12}).$$

Theorem 3. *A given object-image pair are related, i.e. $K^1 \subset L^2$, if and only if the following equations are satisfied:*

$$p_1 M_{23} - p_2 M_{13} + p_3 M_{12} = 0$$

$$p_1 M_{24} - p_2 M_{14} + p_4 M_{12} = 0$$

$$p_1 M_{34} - p_3 M_{14} + p_4 M_{13} = 0$$

$$p_2 M_{34} - p_3 M_{24} + p_4 M_{23} = 0.$$

These equations, which are bihomogeneous of bidegree (1,1) in the object-image coordinates, together with the Plücker relation

$$M_{12} M_{34} - M_{13} M_{24} + M_{14} M_{23} = 0$$

cut out the subvariety $Z^5 \subset X^3 \times Y^4 \subset \mathbf{P}^3 \times \mathbf{P}^5$ which gives our correspondence.

Proof: This is just a computation in multilinear algebra. Consider the vector

$$\vec{v} = p_1 \vec{e}_1 + p_2 \vec{e}_2 + p_3 \vec{e}_3 + p_4 \vec{e}_4 \in H^4$$

which spans K^1 and the bi-vector

$$\vec{w} = M_{12} \vec{e}_1 \wedge \vec{e}_2 + M_{13} \vec{e}_1 \wedge \vec{e}_3 + M_{14} \vec{e}_1 \wedge \vec{e}_4 + M_{23} \vec{e}_2 \wedge \vec{e}_3$$
$$+ M_{24} \vec{e}_2 \wedge \vec{e}_4 + M_{34} \vec{e}_3 \wedge \vec{e}_4 \in \Lambda^2 H^4$$

which "spans" L^2. We will have $K^1 \subset L^2$ if and only if $\vec{v} \in L^2$. This is equivalent to $\vec{w} \wedge \vec{v} = \vec{0} \in \Lambda^3 H^4$. We have

$$\vec{w} \wedge \vec{v} = (p_1 M_{23} - p_2 M_{13} + p_3 M_{12}) \vec{e}_1 \wedge \vec{e}_2 \wedge \vec{e}_3 +$$
$$(p_1 M_{24} - p_2 M_{14} + p_4 M_{12}) \vec{e}_1 \wedge \vec{e}_2 \wedge \vec{e}_4 +$$
$$(p_1 M_{34} - p_3 M_{14} + p_4 M_{13}) \vec{e}_1 \wedge \vec{e}_3 \wedge \vec{e}_4 +$$
$$(p_2 M_{34} - p_3 M_{24} + p_4 M_{23}) \vec{e}_2 \wedge \vec{e}_3 \wedge \vec{e}_4$$

and the result follows. (Note: $\vec{e}_1, \vec{e}_2, \vec{e}_3, \vec{e}_4$ is the previously discussed basis of H^4.) □

To summarize in another form:

Theorem 4. *Let $P_i = (x_i, y_i, z_i)$, $i = 0, \ldots, 4$ be an ordered five-tuple of points in space thought of as a feature set of a 3-D object and let $Q_j = (u_j, v_j)$ be an ordered five-tuple of points in the plane thought of as the features in an image. Assume only that the P_i are not all coplanar and that the Q_j are not all collinear. Define*

$$p_1 = \det \begin{pmatrix} x_0 & x_2 & x_3 & x_4 \\ y_0 & y_2 & y_3 & y_4 \\ z_0 & z_2 & z_3 & z_4 \\ 1 & 1 & 1 & 1 \end{pmatrix}$$

$$p_2 = -\det \begin{pmatrix} x_0 & x_1 & x_3 & x_4 \\ y_0 & y_1 & y_3 & y_4 \\ z_0 & z_1 & z_3 & z_4 \\ 1 & 1 & 1 & 1 \end{pmatrix}$$

$$p_3 = \det \begin{pmatrix} x_0 & x_1 & x_2 & x_4 \\ y_0 & y_1 & y_2 & y_4 \\ z_0 & z_1 & z_2 & z_4 \\ 1 & 1 & 1 & 1 \end{pmatrix}$$

$$p_4 = -\det \begin{pmatrix} x_0 & x_1 & x_2 & x_3 \\ y_0 & y_1 & y_2 & y_3 \\ z_0 & z_1 & z_2 & z_3 \\ 1 & 1 & 1 & 1 \end{pmatrix}$$

and

$$M_{12} = \det \begin{pmatrix} u_0 & u_3 & u_4 \\ v_0 & v_3 & v_4 \\ 1 & 1 & 1 \end{pmatrix}$$

$$M_{13} = -\det \begin{pmatrix} u_0 & u_2 & u_4 \\ v_0 & v_2 & v_4 \\ 1 & 1 & 1 \end{pmatrix}$$

$$M_{14} = \det \begin{pmatrix} u_0 & u_2 & u_3 \\ v_0 & v_2 & v_3 \\ 1 & 1 & 1 \end{pmatrix}$$

$$M_{23} = \det \begin{pmatrix} u_0 & u_1 & u_4 \\ v_0 & v_1 & v_4 \\ 1 & 1 & 1 \end{pmatrix}$$

$$M_{24} = -\det \begin{pmatrix} u_0 & u_1 & u_3 \\ v_0 & v_1 & v_3 \\ 1 & 1 & 1 \end{pmatrix}$$

$$M_{34} = \det \begin{pmatrix} u_0 & u_1 & u_2 \\ v_0 & v_1 & v_2 \\ 1 & 1 & 1 \end{pmatrix}.$$

Then the Q_j, $j = 0, \ldots, 4$ are the images of the P_i, $i = 0, \ldots, 4$ under a generalized weak perspective transformation if and only if

$$p_1 M_{23} - p_2 M_{13} + p_3 M_{12} = 0$$

$$p_1 M_{24} - p_2 M_{14} + p_4 M_{12} = 0$$

$$p_1 M_{34} - p_3 M_{14} + p_4 M_{13} = 0$$

$$p_2 M_{34} - p_3 M_{24} + p_4 M_{23} = 0.$$

These polynomials are bihomogeneous of bidegree (3,2) in the variables x_i, y_i, z_i and u_j, v_j respectively. □

For a fixed object, the set $\sum_{K^1}^2$ of all images it produces is a projective plane $\mathbf{P}^2 \subset Gr(2,4) \subset \mathbf{P}^5$ described by the above equations and the equation $M_{12}M_{34} - M_{13}M_{24} + M_{14}M_{23} = 0$. On the other hand, for a fixed image, the set $\sum_{L^2}^1$ of all objects which can produce that image is a projective line $\mathbf{P}^1 \subset \mathbf{P}^3$. To see these facts, we need to work in an open subset of $\mathbf{P}^3 \times \mathbf{P}^5$.

We choose the open set $\mathbf{R}^3 \times \mathbf{R}^5$ given by $p_4 \neq 0$ and $M_{34} \neq 0$. (This is equivalent to assuming P_0, P_1, P_2, P_3 are not coplanar and that Q_0, Q_1, Q_2 are not collinear.) On this open set we can normalize our projective coordinates on $\mathbf{P}^3$ and $\mathbf{P}^5$ so that $p_4 = -1$ and $M_{34} = 1$. This yields coordinates

$$(\widetilde{p}_1 : \widetilde{p}_2 : \widetilde{p}_3 : -1) \quad \text{and} \quad (\widetilde{q}_1 \widetilde{\widetilde{q}}_2 - \widetilde{\widetilde{q}}_1 \widetilde{q}_2 : \widetilde{\widetilde{q}}_1 : -\widetilde{q}_1 : \widetilde{\widetilde{q}}_2 : -\widetilde{q}_2 : 1)$$

where $\widetilde{p}_i, \widetilde{q}_j$ and $\widetilde{\widetilde{q}}_k$ have been previously defined.

In terms of these variables our equations become:

A) $$\widetilde{p}_1 \widetilde{\widetilde{q}}_2 - \widetilde{p}_2 \widetilde{\widetilde{q}}_1 + \widetilde{p}_3 (\widetilde{q}_1 \widetilde{\widetilde{q}}_2 - \widetilde{\widetilde{q}}_1 \widetilde{q}_2) = 0$$

B) $$-\widetilde{p}_1\widetilde{q}_2 + \widetilde{p}_2\widetilde{q}_1(\widetilde{q}_1\widetilde{\widetilde{q}}_2 - \widetilde{\widetilde{q}}_1\widetilde{q}_2) = 0$$

C) $$\widetilde{p}_1 + \widetilde{p}_3\widetilde{q}_1 - \widetilde{\widetilde{q}}_1 = 0$$

D) $$\widetilde{p}_2 + \widetilde{p}_3\widetilde{q}_2 - \widetilde{\widetilde{q}}_2 = 0.$$

However the first two equations are consequences of the last two. For example, equation A) is $\widetilde{\widetilde{q}}_2$ times equation C) minus $\widetilde{\widetilde{q}}_1$ times equation D). Equations C) and D) are just (1) and (2) above.

These equations can be generalized to more points (we have removed the superscripts and reindexed the q's):

$$\begin{aligned} p_1 + q_1p_3 - q_3 &= 0 \\ p_2 + q_2p_3 - q_4 &= 0 \\ p_4 + q_1p_6 - q_5 &= 0 \\ p_5 + q_2p_3 - q_6 &= 0 \\ &\vdots \end{aligned}$$

Here $(p_1, p_2, p_3), (p_4, p_5, p_6)$, etc. are the affine invariants of our 3D point set. These invariants are obtained by moving the first four points to (0,0,0), (1,0,0), (0,1,0), and (0,0,1) respectively, so that (p_1, p_2, p_3) are the coordinates of the fifth point, (p_4, p_5, p_6) are the coordinates of the sixth point, etc. Also (q_1, q_2), (q_3, p_4), (q_5, q_6), etc. are the affine invariants of our 2D points set. These are similarly obtained by moving the first three points to (0,0), (1,0) and (0,1) respectively, so that (q_1, q_2), (q_3, q_4), (q_5, q_6), etc. are the coordinates of the fourth, fifth, sixth, etc., points.

Example: In our example where $P_0 = (1, 1, 2)$, etc. and $Q_0 = (-5/6, 3/2)$, etc. We have $p_1 = 1$, $p_2 = -10$, $p_3 = -9$, $p_4 = 3$, $M_{12} = 25/108$, $M_{13} = 5/27$, $M_{14} = -25/216$, $M_{23} = 25/108$, $M_{24} = 25/54$, $M_{34} = 35/72$, and one easily verifies that these satisfy our equations. Note that

$$\begin{aligned} \widetilde{p}_1 &= -\frac{p_1}{p_4} = -1/3 \\ \widetilde{p}_2 &= -\frac{p_2}{p_4} = 10/3 \\ \widetilde{p}_3 &= -p_3/p_4 = 3 \\ \widetilde{q}_1 &= -M_{14}/M_{34} = 5/21 \end{aligned}$$

$$\begin{aligned}
\widetilde{q}_2 &= -M_{24}/M_{34} = -\frac{20}{21} \\
\widetilde{\widetilde{q}}_1 &= M_{13}/M_{34} = \frac{8}{21} \\
\widetilde{\widetilde{q}}_2 &= M_{23}/M_{24} = \frac{10}{21}.
\end{aligned}$$

(Watch out for the signs in the definition of M_{24} and M_{13}.) Also

$$\widetilde{q}_1\widetilde{\widetilde{q}}_2 - \widetilde{\widetilde{q}}_1\widetilde{q}_2 = \frac{10}{21} = \frac{M_{12}}{M_{34}}.$$

§5. Lines

We now wish to consider m lines $\ell_1, \ldots, \ell_m$ in addition to our n-points $P_0, \ldots, P_{n-1}$. As before, we assume $n \geq 4$. For the moment, we also assume that P_0, P_1, P_2, P_3 are not coplanar. Each line ℓ_j can be represented by two linear equations, $a_j x + b_j y + c_j z + d_j = 0$ and $e_j x + f_j y + g_j z + h_j = 0$, which we put into a 2×4 array:

$$\begin{pmatrix} a_j & b_j & c_j & d_j \\ e_j & f_j & g_j & h_j \end{pmatrix}.$$

These equations are not unique. Any other set of two equations defining ℓ_j can be obtained from the above set by taking linear combinations. This means that the 2×4 array above is determined only up to left multiplication by an invertible 2×2 matrix:

$$\begin{pmatrix} r_j & s_j \\ t_j & u_j \end{pmatrix} \quad \begin{pmatrix} a_j & b_j & c_j & d_j \\ e_j & f_j & g_j & h_j \end{pmatrix},$$

where $r_j u_j - s_j t_j \neq 0$. To avoid this ambiguity, one usually takes the so-called "line coordinates" of ℓ_j which are the homogeneous coordinates of a point in $\mathbf{P}^5$:

$$\begin{aligned}(m_{12} : m_{13} : m_{14} : m_{23} : m_{24} : m_{34}) = \Bigg(\det \begin{pmatrix} a_j & b_j \\ e_j & f_j \end{pmatrix} : \det \begin{pmatrix} a_j & c_j \\ e_j & g_j \end{pmatrix} : \det \begin{pmatrix} a_j & d_j \\ f_j & g_j \end{pmatrix} \\ : \det \begin{pmatrix} b_j & c_j \\ e_j & h_j \end{pmatrix} : \det \begin{pmatrix} b_j & d_j \\ f_j & h_j \end{pmatrix} : \det \begin{pmatrix} c_j & d_j \\ g_j & h_j \end{pmatrix}\Bigg).\end{aligned}$$

This point always lies in the quadric hypersurface defined by $m_{12}m_{34} - m_{13}m_{24} + m_{14}m_{23} = 0$. This hypersurface is the Grassmannian $Gr_{\mathbf{R}}(2,4)$ embedded via the Plücker embedding in $\mathbf{P}^5$. Recall that a dense open subset of the four dimensional manifold $Gr_{\mathbf{R}}(2,4)$ parameterizes the set of all lines in $\mathbf{R}^3$. Specifically, lines in $\mathbf{R}^3$ are parameterized by $Gr_{\mathbf{R}}(2,4)$ minus the plane cut out by $m_{12} = 0$, $m_{13} = 0$, and $m_{23} = 0$.

The group of affine transformations of $\mathbf{R}^3$, $AFF(3, \mathbf{R})$, acts on lines. In particular, if

$$M = \begin{pmatrix} a & b & c & \xi_1 \\ d & e & f & \xi_2 \\ g & h & k & \xi_3 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

is an affine transformation, then M acts on the equations of the line ℓ_j by:

$$\begin{pmatrix} a_j & b_j & c_j & d_j \\ e_j & f_j & g_j & h_j \end{pmatrix} M^{-1}.$$

We obtain affine invariants for the geometric configuration consisting of $P_0, \ldots, P_{n-1}$, $\ell_1, \ldots, \ell_m$ by moving P_0, P_1, P_2, P_3 to (0,0,0), (1,0,0), (0,1,0), (0,0,1) respectively. If we call the matrix that does this M, then

$$M^{-1} = \begin{pmatrix} x_1 - x_0 & x_2 - x_0 & x_3 - x_0 & x_0 \\ y_1 - y_0 & y_2 - y_0 & y_3 - y_0 & y_0 \\ z_1 - z_0 & z_2 - z_0 & z_3 - z_0 & z_0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

and the line coordinates (actually the five ratios obtained by dividing by one particular line coordinate) of

$$\begin{pmatrix} a_j & b_j & c_j & d_j \\ e_j & f_j & g_j & h_j \end{pmatrix} M^{-1}$$

are the fundamental invariants. Essentially each line yields 4 invariants because there are five ratios, but one relation which expresses one ratio in terms of the other four. Thus we have a total of $3n + 4m - 12$ fundamental invariants. One can easily show that any affine invariant of n points and m lines will be a rational function of the $3n + 4m - 12$ fundamental invariants described above.

Thus, after moving P_0, P_1, P_2, P_3 to standard position, we have the ratios of the line coordinates providing four independent invariants for each line. Our keys $K_j^2 \subset \mathbf{R}^4$ for the lines $\ell_1, \ldots, \ell_m$ will be the planes spanned by the rows of the products

$$\begin{pmatrix} a_j & b_j & c_j & d_j \\ e_j & f_j & g_j & h_j \end{pmatrix} \begin{pmatrix} x_1 - x_0 & x_2 - x_0 & x_3 - x_0 & x_0 \\ y_1 - y_0 & y_2 - y_0 & y_3 - y_0 & y_0 \\ z_1 - z_0 & z_2 - z_0 & z_3 - z_0 & z_0 \\ 0 & 0 & 0 & 1 \end{pmatrix}, \quad j = 1, \ldots, n.$$

As we have previously described, each K_j^2 carries the invariants of the particular line ℓ_j involved, i.e. knowledge of K_j^2 is equivalent to knowing the invariants contributed by the line.

Now suppose that $Q_0 = (u_0, v_0), Q_1 = (u_1, v_1)$, $Q_2 = (u_2, v_2)$, and $Q_3 = (u_3, v_3)$ are the images of P_0, P_1, P_2, P_3 and that $r_j u + m_j v + n_j = 0$ is the image of the line ℓ_j. We consider the matrix product

$$(5) \qquad \begin{pmatrix} r_j & m_j & 0 & n_j \\ * & * & * & * \end{pmatrix} \begin{pmatrix} u_1 - u_0 & u_2 - u_0 & u_3 - u_0 & u_0 \\ v_1 - v_0 & v_2 - v_0 & v_3 - v_0 & v_0 \\ * & * & * & * \\ 0 & 0 & 0 & 1 \end{pmatrix} = \begin{pmatrix} \alpha_j & \beta_j & \gamma_j & \delta_j \\ * & * & * & * \end{pmatrix}.$$

The lines $L_j^1 \subset \mathbf{R}^4$ spanned by $(\alpha_j, \beta_j, \gamma_j, \delta_j)$ will be our lookups.

Theorem 5. *The spaces L_j^1 and K_j^2 are invariant under affine transformation in 2-D and 3-D respectively and for each j, $L_j^1 \subset K_j^2 \subset \mathbf{R}^4$.*

Proof: Invariance is clear. Also, it is easy to see that the entire process of taking a weak-perspective transformation and following it by an affine change of coordinates in the image plane is equivalent to applying an affine transformation followed by orthogonal projection. Thus, with no loss of generality, we may assume that P_i orthogonally projects to Q_i for $i = 0, 1, 2, 3$, i.e. that $x_i = u_i$ and $y_i = v_i$ for $i = 0, 1, 2, 3$, and that ℓ_j projects orthogonally to the line $r_j u + m_j v + n_i = 0$. This means that the rows of

$$\begin{pmatrix} \alpha_j & \beta_j & \gamma_j & \delta_j \\ * & * & * & * \end{pmatrix}$$

in (5) span the plane K^2 and that $L^1 \subset K^2$. $\square$

In fact, if $L_j^1 \not\subset K_j^2$ for some j, then the image could not have been produced by the given 3-D arrangement of points $P_0, \ldots, P_{n-1}$ and lines $\ell_1, \ldots, \ell_j, \ldots \ \ell_m$. For the recognition or indexing problem, we will need to find all objects with keys K_j^2 that contain the lookups L_j^1 respectively. Again, we can interpret each key K_j^2 as a point in the Grassmannian of two planes in four space, $Gr_{\mathbf{R}}(2, 4)$. Each lookup L_j^1 defines a Schubert cycle $\Psi_{L_j^1} \subset Gr_{\mathbf{R}}(2, 4)$ consisting of all $[K^2] \in Gr_{\mathbf{R}}(2, 4)$ such that $L_j^1 \subset K^2$. It is well-known that $\Psi_{L_j^1}$ is a linear $\mathbf{P}^2$, i.e. a plane in $Gr_{\mathbf{R}}(2, 4) \subset \mathbf{P}^5$ (see Harris [3] pg. 67).

Notice that when line ℓ_j is occluded, we can take $L_j^1 = \{0\}$ so that no constraint is imposed. In effect, we just ignore the missing line.

Example: We continue with the example above where $P_0 = (1, 1, 2)$, $P_1 = (0, -2, 1)$, $P_2 = (3, 1, 1)$, $P_3 = (-1, 0, 3)$, and $P_4 = (2, -1, 2)$. Let's add the line ℓ_1 defined by $x + y + z - 1 = 0$ and $y = 0$. To calculate the invariants and our key, we look at

$$\begin{pmatrix} 1 & 1 & 1 & -1 \\ 0 & 1 & 0 & 0 \end{pmatrix} \begin{pmatrix} -1 & 2 & -2 & 1 \\ -3 & 0 & -1 & 1 \\ -1 & -1 & 1 & 2 \\ 0 & 0 & 0 & 1 \end{pmatrix} = \begin{pmatrix} -5 & 1 & -2 & 3 \\ -3 & 0 & -1 & 1 \end{pmatrix}.$$

Our key K_1^2 is the plane spanned by the vectors $(-5, 1, -2, 3)$ and $(-3, 0, -1, 1)$.

If we apply the rigid motion from the example above to establish our view, our line becomes

$$\begin{pmatrix} 1 & 1 & 1 & -1 \\ 0 & 1 & 0 & 0 \end{pmatrix} \begin{pmatrix} \frac{2}{3} & \frac{2}{3} & \frac{1}{2} & -\frac{4}{3} \\ -\frac{1}{3} & \frac{2}{3} & -\frac{2}{3} & \frac{5}{3} \\ -\frac{2}{3} & \frac{1}{3} & \frac{2}{3} & -\frac{2}{3} \\ 0 & 0 & 0 & 1 \end{pmatrix} = \begin{pmatrix} -\frac{1}{3} & \frac{5}{3} & \frac{1}{3} & -\frac{4}{3} \\ -\frac{1}{3} & \frac{2}{3} & -\frac{2}{3} & \frac{5}{3} \end{pmatrix}.$$

It then projects orthogonally to the line $-x + 4y - 1 = 0$ which, under scaling by $\frac{1}{12}$, becomes

$$-12u' + 48v' - 1 = 0.$$

The choice of coordinates in the image plane gives:

$$(-12, 48, -1) \begin{pmatrix} 3 & 1 & -1 \\ -1 & 3 & 1 \\ 0 & 0 & 1 \end{pmatrix}^{-1} =$$

$$(-12, 48, -1) \begin{pmatrix} \frac{3}{10} & -\frac{1}{10} & \frac{4}{10} \\ \frac{1}{10} & \frac{3}{10} & -\frac{2}{10} \\ 0 & 0 & 1 \end{pmatrix} = (1.2, 15.6, -15.4).$$

Thus in the image plane our line is $1.2u + 15.6v - 15.4 = 0$.

To construct our lookup, we form the matrix product

$$\begin{pmatrix} 1.2 & 15.6 & 0 & -15.4 \\ * & * & * & * \end{pmatrix} \begin{pmatrix} 0 & \frac{7}{12} & -\frac{5}{9} & -\frac{5}{6} \\ -\frac{5}{6} & \frac{1}{12} & -\frac{5}{18} & \frac{3}{2} \\ * & * & * & * \\ 0 & 0 & 0 & 1 \end{pmatrix} = \begin{pmatrix} -13 & 2 & -5 & 7 \\ * & * & * & * \end{pmatrix}$$

and take the line L_1^1 spanned by $(-13, 2, -5, 7)$ in $\mathbf{R}^4$.

Notice that $L_1^1 \subset K_1^2$ because the vector $(-13, 2, -5, 7)$ is a linear combination of $(-5, 1, -2, 3)$ and $(-3, 0, -1, 1)$.

We note that the general position assumption, namely that P_0, P_1, P_2, P_3 not be coplanar, can be relaxed to the assumption that $P_0, \ldots, P_{n-1}$ are not all coplanar. In that case, our keys K_j^2 can be taken to be the span of the two rows of

$$\begin{pmatrix} a_j & b_j & c_j & d_j \\ e_j & f_j & g_j & h_j \end{pmatrix} \begin{pmatrix} x_1 - x_0 & & x_{n-1} - x_0 & x_0 \\ y_1 - y_0 & \cdots & y_{n-1} - y_0 & y_0 \\ z_1 - z_0 & & z_{n-1} - z_0 & z_0 \\ 0 & & 0 & 1 \end{pmatrix}$$

and our lookups L_j^1 can be taken to be the span of the first row of the matrix product

$$\begin{pmatrix} r_j & m_j & 0 & n_j \\ * & * & * & * \end{pmatrix} \quad \begin{pmatrix} u_1 - u_0 & & u_{n-1} - u_0 & u_0 \\ v_1 - v_0 & \dots & v_{n-1} - v_0 & v_0 \\ * & & * & * \\ 0 & & 0 & 1 \end{pmatrix}.$$

We mention that when the points are not available, for example if we are dealing with only lines, then other methods can be used. These are discussed in Asmuth, Stiller and Wan [8].

§6. Conclusions

Using the theory of correspondences from algebraic geometry, we developed methods to relate 3-D objects to 2-D images and vice versa. In effect, we have provided a very general framework for the use of geometric invariants in image recognition. Our techniques yield systems of polynomial equations in both the 3-D invariants of various features on an object and the 2-D invariants of features in an image which will be satisfied if and only if the object can produce the image up to affine transformations of both the object and the image. The case of projective invariants and perspective projection have been worked out for point and line features and will be explained in a forthcoming paper [8].

The subspace incidence relation $K^{n-4} \subset L^{n-3} \subset H^{n-1} \subset \mathbf{R}^n$ can be framed in terms of some very natural metrics that measure when K^{n-4} is "close" to being included in L^{n-3}. These metrics arise in the context of Grassmannians and would seem to be the most natural and robust measures of the "distances" between two objects, between two images, and between an object and an image. Moreover, the explicit equations we derive make possible an effective error analysis.

In addition, many well-known results are easily derived from within our framework. For example, the non-existence of general view invariants (see [1]) follows from the fact that given two ordered sets of n-points in $\mathbf{R}^3$, i.e. K^{n-4} and $\widetilde{K}^{n-4}$, we can find a sequence

$$K^{n-4} = K_0^{n-4}, K_1^{n-4}, K_2^{n-4}, \ldots, K_s^{n-4} = \widetilde{K}^{n-4}$$

of subspaces with the span of K_i^{n-4} and K_{i+1}^{n-4} being a linear space L_i^{n-3} of dimension $n-3$. Since K_i^{n-4} and K_{i+1}^{n-4} have a common image, L_i^{n-3}, a general view invariant would take the same value on images of these objects. It then follows that any general view invariant is a constant.

Finally, the applications we have considered are to single view recognition and to indexing image databases for content based retrieval (see [7]). The actual implementation of these ideas has been carried out in the case of indexing an image database for content based retrieval. Details on the indexing method and on the performance and error analysis can be found in [7].

Appendix 1. $Gr(2,4)$ **and orthogonal planes.**

Given $\sigma \subset \mathbf{R}^4$ spanned by $\overrightarrow{a} = (a_1, a_2, a_3, a_4)^T$ and $\overrightarrow{b} = (b_1, b_2, b_3, b_4)^T$ with rank $\begin{pmatrix} a_1 & a_2 & a_3 & a_4 \\ b_1 & b_2 & b_3 & b_4 \end{pmatrix} = 2$, we get Plücker coordinates for L^2, namely $N_{ij} = \det \begin{pmatrix} a_i & a_j \\ b_i & b_j \end{pmatrix}$, $1 \le i < j \le 4$, with

$$(N_{12} : N_{13} : N_{14} : N_{23} : N_{24} : N_{34}) \in \mathbf{P}^5$$

$$N_{12}N_{34} - N_{13}N_{24} + N_{14}N_{23} = 0.$$

Let $\sigma^\perp$ be the plane which is the kernel of $\begin{pmatrix} a_1 & a_2 & a_3 & a_4 \\ b_1 & b_2 & b_3 & b_4 \end{pmatrix}$ as a linear transformation from $\mathbf{R}^4$ and $\mathbf{R}$ and suppose $\sigma^\perp$ has Plücker coordinates $(M_{12}, \ldots, M_{34}) \in \mathbf{P}^5$.

Proposition.

$$M_{12} = N_{34}$$

$$M_{13} = -N_{24}$$

$$M_{14} = N_{23}$$

$$M_{23} = N_{14}$$

$$M_{24} = -N_{13}$$

$$M_{34} = N_{12}$$

up to scalar.

Proof. Suppose $N_{12} \neq 0$ then σ^2 is spanned by

$$\begin{pmatrix} 1 & 0 & r & s \\ 0 & 1 & t & u \end{pmatrix} = \begin{pmatrix} a_1 & a_2 \\ b_1 & b_2 \end{pmatrix}^{-1} \begin{pmatrix} a_1 & a_2 & a_3 & a_4 \\ b_1 & b_2 & b_3 & b_4 \end{pmatrix}.$$

In particular

$$\begin{aligned} \begin{pmatrix} r & s \\ t & u \end{pmatrix} &= \frac{1}{N_{12}} \begin{pmatrix} b_2 & -a_2 \\ -b_1 & a_1 \end{pmatrix} \begin{pmatrix} a_3 & a_4 \\ b_3 & b_4 \end{pmatrix} \\ &= \frac{1}{N_{12}} \begin{pmatrix} a_3 b_2 - a_2 b_3 & a_4 b_2 - a_2 b_4 \\ -b_1 a_3 + a_1 b_3 & -b_1 a_4 + a_1 b_4 \end{pmatrix} \\ &= \frac{1}{N_{12}} \begin{pmatrix} -N_{23} & -N_{24} \\ N_{13} & N_{14} \end{pmatrix}. \end{aligned}$$

Also $\frac{N_{34}}{N_{12}} = \det \begin{pmatrix} -N_{23}/N_{12} & -N_{24}/N_{12} \\ N_{13}/N_{12} & N_{14}/N_{12} \end{pmatrix} = \det \begin{pmatrix} r & s \\ t & u \end{pmatrix} = ru - st.$

The kernel $\sigma^\perp$ contains vectors such that

$$\begin{pmatrix} 1 & 0 & r & s \\ 0 & 1 & t & u \end{pmatrix} \begin{pmatrix} x \\ y \\ z \\ w \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}.$$

This yields

$$x = -rz - sw$$

$$y = -tz - uw$$

so that $\sigma^{\perp}$ is spanned by the rows of

$$\begin{pmatrix} -r & -t & 1 & 0 \\ -s & -u & 0 & 1 \end{pmatrix}$$

¿From this we get the Plücker coordinates

$$\begin{aligned}(M_{12}, \ldots, M_{34}) &= (ru - st, s, -r, u, -t, 1) \\ &= \left(\frac{N_{34}}{N_{12}}, -\frac{N_{24}}{N_{12}}, \frac{N_{23}}{N_{12}}, \frac{N_{14}}{N_{12}}, -\frac{N_{13}}{N_{12}}, 1 \right).\end{aligned}$$

The results follows. □

An alternative proof can be given using standard multilinear algebra. We ask for the conditions under which a vector

$$\vec{v} = x\,\vec{e}_1 + y\,\vec{e}_2 + z\,\vec{e}_3 + w\,\vec{e}_4$$

lies in the plane defined by

$$\begin{aligned}\vec{w} = {} & M_{34}\,\vec{e}_1 \wedge \vec{e}_2 - M_{24}\,\vec{e}_1 \wedge \vec{e}_3 + M_{23}\,\vec{e}_1 \wedge \vec{e}_4 \\ & + M_{14}\,\vec{e}_2 \wedge \vec{e}_3 - M_{13}\,\vec{e}_2 \wedge \vec{e}_4 + M_{12}\,\vec{e}_3 \wedge \vec{e}_4\end{aligned}$$

where $\vec{e}_1, \vec{e}_2, \vec{e}_3, \vec{e}_4$ is a basis of our four dimensional vector space. The condition is clearly

$$\vec{w} \wedge \vec{v} = 0$$

or

$$\begin{aligned}\vec{0} = {} & xM_{14} + yM_{24} + zM_{34} && \vec{e}_1 \wedge \vec{e}_2 \wedge \vec{e}_3 \\ & -xM_{13} - yM_{23} + wM_{34} && \vec{e}_1 \wedge \vec{e}_2 \wedge \vec{e}_4 \\ & xM_{12} - zM_{23} - wM_{24} && \vec{e}_1 \wedge \vec{e}_3 \wedge \vec{e}_4 \\ & yM_{12} + zM_{13} + wM_{14} && \vec{e}_2 \wedge \vec{e}_3 \wedge \vec{e}_4 .\end{aligned}$$

These conditions are equivalent to

$$\begin{pmatrix} a_1 & a_2 & a_3 & a_4 \\ b_1 & b_2 & b_3 & b_4 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \\ w \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}.$$

For example

$$0 = xM_{14} + yM_{24} + zM_{34} = (b_4(a_1, a_2, a_3, a_4) - a_4(b_1, b_2, b_3, b_4)) \begin{pmatrix} x \\ y \\ z \\ w \end{pmatrix}.$$

Ones sees that the result follows. $\square$

Notice that when $\vec{e}_1, \vec{e}_2, \vec{e}_3, \vec{e}_4$ is an orthonormal basis for H^4 then $\sigma^\perp$ will be the orthogonal complement of $\sigma \subset H^4$. In the application, we do *not* use an orthonormal basis for $H^4 \cong \mathbf{R}^4$, so $\sigma^\perp$ is *not* the orthogonal complement in $H^4 \subset \mathbf{R}^5$, but is the kernel of the appropriate transformation.

Bibliography

[1] Burns, J., Richard S. Weiss and Edward M. Riseman, "The Non-Existence of General-Case View-Invariants," in *Geometric Invariance in Computer Vision*, J.L. Mundy and Andrew Zisserman, eds., MIT Press, 1992.

[2] Clemens, David T., and David W. Jacobs, "Space and Time Bounds on Indexing 3-D Models from 2-D Images," *IEEE Transactions PAMI*, v. 13, n. 10 October 1991.

[3] Harris, "Algebraic Geometry," Graduate Text in Mathematics 133, Springer-Verlag, 1992.

[4] Hartshorne, "Algebraic Geometry," Graduate Text in Mathematics 52, Springer-Verlag, 1977.

[5] Griffiths and Harris, "Principles of Algebraic Geometry," John Wiley and Sons, Inc., 1978.

[6] Mumford, "Geometric Invariant Theory," Springer-Verlag, 1965.

[7] Asmuth, C.A., P.F. Stiller and C.S. Wan, "Progress Report for Project: Indexing Image Databases for Content-Based Retrieval," Internal Report, David Sarnoff Research Center, 1994.

[8] Asmuth, C.A., P.F. Stiller, and C.S. Wan, "A General Theory of Single View Recognition – the Projective Case," in preparation.

[9] Weiss, Issac, "Geometric Invariants and Object Recognition," *International Journal of Computer Vision*, 10.3, 207-231, 1993.

[10] Jacobs, "Matching 3-D Models to 2-D Images," preprint.