Dualizing Scene Reconstruction Algorithms

Richard Hartley and Gilles Debunne

GE–CRD, Schenectady, NY
ENSIMAG, Grenoble.

Abstract

It has been known since the work of Carlsson and Weinshall that there is a du-
alization principle that allows one to interchange the role of points being viewed
by several cameras and the camera centres themselves. In principle this implies
the possibility of dualizing projective reconstruction algorithms to obtain new
algorithms. In this paper, this theme is developed at a theoretical and algorith-
mic level. The nature of the duality mapping is explored and its application
to reconstruction ambiguity is discussed. An explicit method for dualizing any
projective reconstruction algorithm is given. At the practical implementation
level, however, it is shown that there are difficulties which have so far defeated
successful application of this dualization method to produce working algorithms.

1 Introduction

The theory and practice of projective and metric reconstruction from uncali-
brated and semi calibrated views has reached a level of maturity in recent years
that excellent results may now be achieved. Papers presented at this workshop
and reported in this volume show the high quality of reconstruction that is now
possible.

In particular, it would appear that many of the problems of reconstruction
have now reached a level where one may claim that they are solved. Such prob-
lems include

1. Computation of the multifocal tensors, particularly the fundamental matrix
   and trifocal tensors (the quadrifocal tensor having not received so much
   attention).
2. Extraction of the camera matrices from these tensors, and subsequent pro-
   jective reconstruction from two and three views.

Other significant successes have been achieved, though there may be more to
learn about these problems.

1. Application of bundle adjustement to solve more general reconstruction
   problems.
2. Metric (Euclidean) reconstruction given minimal assumptions on the camera
   matrices.

3. Automatic detection of correspondences in image sequences, and elimination of outliers and false matches using the multifocal tensor relationships.

In other areas the last word has clearly not been written. Notably, there is not any single satisfactory algorithm for projective reconstruction from several views. Many methods have been tried : iterative methods, methods based on tacking together reconstructions from small numbers of views, or factorization-based algorithms, which need arbitrary guesses at depth.

This paper discusses a technique that, although known, seems not to have received as much attention as may be warranted. The method based on a dualization principle expounded by Carlsson and also Weinshall ([?,?]) can in principle transform the problem of projective reconstruction from long image sequences into the problem of projective reconstruction from small numbers of views, for which (as claimed above) the reconstruction problem is essentially solved. It is shown that although this duality theoretically gives rise to the desired multiple-view algorithms, in reality there are practical difficulties. In this paper, the problem of how to obtain working algorithms from this method is not solved. The purpose is to highlight the fascinating properties of the duality method, here called Carlsson duality, with the hope of awaking enough interest to lead to a practical implementation of these methods.

Before we proceed to discuss duality, I claim the privilege of giving an opinion. At this point of maturity, the understanding of the underlying geometrical properties of multi-view vision and the implementation of high-quality geometrical algorithms have outstripped the less mathematically structured tasks of correspondence matching and 3D model building that are essential to building a good system (despite the excellent results achieved and reported at the workshop). In short, we seem to be able to obtain small robust sets of image correspondences and reconstruct these points in 3-space. But how does one find sufficiently many correspondences to build a complete model, and anyway, how does one build a complete 3D model, that is, fill in the gaps between the points ? We can still not do satisfactory automatic reconstruction from complex outdoor scenes (for instance a forest scene) or even indoor scenes such as a room with a jumble of furniture and equipment (such as my office). However, leaving for another day a consideration of these harder problem, we now turn to the main technical subject of this paper.

2 Carlsson Duality

Let $\mathbf{E}_1 = (1, 0, 0, 0)^\top$, $\mathbf{E}_2 = (0, 1, 0, 0)^\top$, $\mathbf{E}_3 = (0, 0, 1, 0)^\top$ and $\mathbf{E}_4 = (0, 0, 0, 1)^\top$ form part of a projective basis for $\mathcal{P}^3$. Similarly, let $\mathbf{e}_1 = (1, 0, 0)^\top$ $\mathbf{e}_2 = (0, 1, 0)^\top$ $\mathbf{e}_3 = (0, 0, 1)^\top$ $\mathbf{e}_4 = (1, 1, 1)^\top$ be a projective basis for the projective image plane $\mathcal{P}^2$.

Now, consider a camera with matrix P. We assume that the camera centre does not sit on any of the axial planes, that is $\mathbf{X} = (\text{X}, \text{Y}, \text{Z}, \text{T})^\top$ and none of the four coordinates is zero. In this case, no three of the points $\text{P}\mathbf{E}_i$ for $i = 1, \ldots, 3$ are collinear in the image. Consequently, one may apply a projective transformation

H to the image so that $\mathbf{e}_i = \mathtt{HPE}_i$. We assume that this has been done, and henceforth denote $\mathtt{HP}$ simply by $\mathtt{P}$. Since $\mathtt{PE}_i = \mathbf{e}_i$, one computes that the form of the matrix $\mathtt{P}$ is

$$\mathtt{P} = \begin{bmatrix} \alpha^{-1} & & & -\delta^{-1} \\ & \beta^{-1} & & -\delta^{-1} \\ & & \gamma^{-1} & -\delta^{-1} \end{bmatrix} \quad . \tag{1}$$

Further, the camera centre is $\mathbf{C} = (\alpha, \beta, \gamma, \delta)^\top$, as one verifies by solving $\mathtt{PC} = \mathbf{0}$. If $\mathbf{C} = (\alpha, \beta, \gamma, \delta)^\top$ is any point in $\mathcal{P}^3$, then the matrix in (1) will be denoted by $\mathtt{P}_\mathbf{C}$.

Now, for any point $\mathbf{X} = (\mathrm{X}, \mathrm{Y}, \mathrm{Z}, \mathrm{T})^\top$ one verifies that

$$\mathtt{P}_\mathbf{C}\,\mathbf{X} = \begin{pmatrix} \alpha^{-1}\mathrm{X} - \delta^{-1}\mathrm{T} \\ \beta^{-1}\mathrm{Y} - \delta^{-1}\mathrm{T} \\ \gamma^{-1}\mathrm{Z} - \delta^{-1}\mathrm{T} \end{pmatrix} \quad . \tag{2}$$

This observation leads to the following definition

Definition 1. *The mapping of $\mathcal{P}^3$ to itself given by*

$$(\mathrm{X}, \mathrm{Y}, \mathrm{Z}, \mathrm{T})^\top \mapsto (\mathrm{YZT}, \mathrm{ZTX}, \mathrm{TXY}, \mathrm{XYZ})^\top$$

will be called the Carlsson map, and will be denoted by Γ. We denote the image of a point $\mathbf{X}$ under Γ by $\overline{\mathbf{X}}$. The image of an object under Γ is sometimes referred to as the dual *object, for reasons that will be seen later.*

The Carlsson map is an example of a *Cremona* transformation. For more information on Cremona transformations, the reader is refered to Semple and Kneebone ([?]).

Note. If none of the coordinates of $\mathbf{X}$ is zero then we may divide $\overline{\mathbf{X}}$ by XYZT. Then Γ is equivalent to $(\mathrm{X}, \mathrm{Y}, \mathrm{Z}, \mathrm{T})^\top \mapsto (\mathrm{X}^{-1}, \mathrm{Y}^{-1}, \mathrm{Z}^{-1}, \mathrm{T}^{-1})^\top$. This is the form of the mapping that we will usually use. In the case where one of the coordinates of $\mathbf{X}$ is zero, then the mapping will be interpreted as in the definition. Note that any point $(0, \mathrm{Y}, \mathrm{Z}, \mathrm{T})^\top$ is mapped to the point $(1, 0, 0, 0)^\top$ by Γ, provided none of the other coordinates is zero. Thus, the mapping is not one-to-one.

If two of the coordinates of $\mathbf{X}$ are zero, then $\overline{\mathbf{X}} = (0, 0, 0, 0)^\top$, which is an undefined point. Thus, Γ is not defined at all points. In fact, there is no way to extend Γ continuously to such points. Note that the points for which the mapping is undefined consists of the lines joining two of the points $\mathbf{E}_i$. We will call the four points $\mathbf{E}_i$ the vertices of the *reference tetrahedron*. The lines joining two vertices are the edges of the tetrahedron, and the planes defined by three vertices are the faces of the reference tetrahedron. As remarked, Γ is undefined on the edges of the reference tetrahedron. As for the faces of the reference tetrahedron, these are the points with a zero coordinate. Consequently (as shown above), each face is mapped by Γ to a single point, namely the opposite vertex of the reference tetrahedron.

Duality

The major importance of the Carlsson map derives from the following formula, which is easily derived from (2).

$$\mathrm{P_C X} = \mathrm{P_{\overline{X}} \overline{C}} \tag{3}$$

Thus, Γ interchanges the rôles of object points and camera centres. Thus, $\mathbf{C}$ acting on $\mathbf{X}$ gives the same result as $\overline{\mathbf{X}}$ acting on $\overline{\mathbf{C}}$. The consequences of this result will be investigated soon. However, first we will investigate the way in which Γ acts on other geometric objects.

Theorem (2.0.1). *The Carlsson map, Γ acts in the following manner :*

1. *It maps a line passing through two general points $\mathbf{X}_0$ and $\mathbf{X}_1$ to the twisted cubic passing through $\overline{\mathbf{X}}_0, \overline{\mathbf{X}}_1$ and the the four reference vertices $\mathbf{E}_1, \ldots, \mathbf{E}_4$.*
2. *It maps a line passing through any of the points $\mathbf{E}_i$ to a line passing through the same $\mathbf{E}_i$. We exclude the lines lying on the face of the reference tetrahedron, since such lines will be mapped to a single point.*
3. *It maps a quadric $\mathbf{Q}$ passing through the four points $\mathbf{E}_i$, $i = 1, \ldots 4$ to a quadric surface (denoted $\overline{\mathbf{Q}}$) passing through the same four points. If $\mathbf{Q}$ is a ruled quadric, then so is $\overline{\mathbf{Q}}$. If $\mathbf{Q}$ is degenerate then so is $\bar{\mathbf{Q}}$.*

Proof. **Part 1.** A line has parametric equation $(\mathrm{X}_0 + a\theta, \mathrm{Y}_0 + b\theta, \mathrm{Z}_0 + c\theta, \mathrm{T}_0 + d\theta)^\top$, and a point on this line is taken by the Carlsson map to the point

$$((\mathrm{Y}_0 + b\theta)(\mathrm{Z}_0 + c\theta)(\mathrm{T}_0 + d\theta), \ldots, (\mathrm{X}_0 + a\theta)(\mathrm{Y}_0 + b\theta)(\mathrm{Z}_0 + c\theta))^\top \ .$$

Thus, the entries of the vector are cubic functions of θ, and the curve is a twisted cubic. Now, setting $\theta = -\mathrm{X}_0/a$, the term $(\mathrm{X}_0 + a\theta)$ vanishes, and the corresponding dual point is $((\mathrm{Y}_0 + b\theta)(\mathrm{Z}_0 + c\theta)(\mathrm{T}_0 + d\theta), 0, 0, 0)^\top \approx (1, 0, 0, 0)^\top$. The first entry is the only one that does not contain $(\mathrm{X}_0 + a\theta)$, and hence the only one that does not vanish. This shows that the reference vertex $\mathbf{E}_1 = (1, 0, 0, 0)^\top$ is on the twisted cubic. By similar arguments, the other points $\mathbf{E}_2, \ldots, \mathbf{E}_4$ lie on the twisted cubic also. Note that a twisted cubic is defined by 6 points, and this twisted cubic is defined by the given 6 points $\mathbf{E}_i, \bar{\mathbf{X}}_0, \bar{\mathbf{X}}_1$ that lie on it, where $\mathbf{X}_0$ and $\mathbf{X}_1$ are any two points defining the line.

Part 2. We prove this for lines passing through the point $\mathbf{E}_1 = (1, 0, 0, 0)^\top$. An analogous proof holds for the other points $\mathbf{E}_i$. Choose another point $\mathbf{X} = (\mathrm{X}, \mathrm{Y}, \mathrm{Z}, \mathrm{T})^\top$ on the line, such that $\mathbf{X}$ does not lie on any face of the reference tetrahedron. Thus $\mathbf{X}$ has no zero coordinate. Points on a line passing through $(1, 0, 0, 0)^\top$ and $\mathbf{X} = (\mathrm{X}, \mathrm{Y}, \mathrm{Z}, \mathrm{T})^\top$ are all of the form $(\alpha, \mathrm{Y}, \mathrm{Z}, \mathrm{T})^\top$ for varying values of α. These points are mapped by the transformation to $(\alpha^{-1}, \mathrm{Y}^{-1}, \mathrm{Z}^{-1}, \mathrm{T}^{-1})^\top$. This represents a line passing through the two points $(1, 0, 0, 0)^\top$ and $\overline{\mathbf{X}} = (\mathrm{X}^{-1}, \mathrm{Y}^{-1}, \mathrm{Z}^{-1}, \mathrm{T}^{-1})^\top$.

Part 3. Since the quadric $\mathbf{Q}$ passes through all the points $\mathbf{E}_i$, the diagonal entries of $\mathbf{Q}$ must all be zero. This means that there are no terms involving a squared coordinate (such as X^2) in the equation for the quadric. Hence the

equation for the quadric contains only mixed terms (such as XY, YZ or XT). Therefore, a point $\mathbf{X} = (\mathrm{X}, \mathrm{Y}, \mathrm{Z}, \mathrm{T})^\top$ lies on the quadric $\mathtt{Q}$ if and only if $a\mathrm{XY} + b\mathrm{XZ} + c\mathrm{XT} + d\mathrm{YZ} + e\mathrm{YT} + f\mathrm{ZT} = 0$. Dividing this equation by XYZT, we obtain $a\mathrm{Z}^{-1}\mathrm{T}^{-1} + b\mathrm{Y}^{-1}\mathrm{T}^{-1} + c\mathrm{Y}^{-1}\mathrm{Z}^{-1} + d\mathrm{X}^{-1}\mathrm{T}^{-1} + e\mathrm{X}^{-1}\mathrm{Z}^{-1} + f\mathrm{X}^{-1}\mathrm{Y}^{-1} = 0$. Since $\overline{\mathbf{X}} = (\mathrm{X}^{-1}, \mathrm{Y}^{-1}, \mathrm{Z}^{-1}, \mathrm{T}^{-1})^\top$, this is a quadratic equation in the entries of $\overline{\mathbf{X}}$. Thus Γ maps quadric to quadric. Specifically, suppose $\mathtt{Q}$ is represented by the matrix

$$\mathtt{Q} = \begin{bmatrix} 0 & a & b & c \\ a & 0 & d & e \\ b & d & 0 & f \\ c & e & f & 0 \end{bmatrix} \quad \text{then} \quad \overline{\mathtt{Q}} = \begin{bmatrix} 0 & f & e & d \\ f & 0 & c & b \\ e & c & 0 & a \\ d & b & a & 0 \end{bmatrix}$$

and $\mathbf{X}^\top \mathtt{Q} \mathbf{X} = 0$ implies $\overline{\mathbf{X}}^\top \overline{\mathtt{Q}} \overline{\mathbf{X}} = 0$. The quadric $\overline{\mathtt{Q}}$ is a ruled quadric, since the generators of $\mathtt{Q}$ passing through the point $\mathbf{E}$ map to straight lines, lying on $\overline{\mathtt{Q}}$. One may further verify that $\det \mathtt{Q} = \det \overline{\mathtt{Q}}$, which implies that if $\mathtt{Q}$ is a non-degerate quadric (that is $\det \mathtt{Q} \neq 0$), then so is $\overline{\mathtt{Q}}$. In this non-degenerate case, if $\mathtt{Q}$ is a hyperboloid of one sheet, then $\det \mathtt{Q} > 0$, from which it follows that $\det \overline{\mathtt{Q}} > 0$. Thus $\overline{\mathtt{Q}}$ is also a hyperboloid of one sheet. $\qquad\square$

We wish to interpret duality equation (3) in a coordinate-free manner. The matrix $\mathtt{P_C}$ has by definition the form given in (1), and maps $\mathbf{E}_i$ to $\mathbf{e}_i$ for $i = 1, \ldots, 4$. The image $\mathtt{P_C}\mathbf{X}$ is may be thought of as a representation of the projection of $\mathbf{X}$ relative to the projective basis $\mathbf{e}_i$ in the image. Alternatively, $\mathtt{P_C}\mathbf{X}$ represents the projective equivalence class of the set of the five rays $\overline{\mathbf{CE}_1}, \ldots, \overline{\mathbf{CE}_4}, \overline{\mathbf{CX}}$. Thus $\mathtt{P_C}\mathbf{X} = \mathtt{P_{C'}}\mathbf{X}'$ if and only if the set of rays from $\mathbf{C}$ to $\mathbf{X}$ and the four vertices of the reference tetrahedron is projectively equivalent to the set of rays from $\mathbf{C}'$ to $\mathbf{X}'$ and the four reference vertices.

The duality principle.

There is nothing special about the four points $\mathbf{E}_1, \ldots, \mathbf{E}_4$ used as vertices of the reference tetrahedron, other than the fact that they are non-coplanar. Given any four non-coplanar points, one may define a projective coordinate system in which these four points are the points $\mathbf{E}_i$ forming part of a projective basis. The Carlsson mapping may then be defined with respect to this coordinate frame. The resulting map is called the Carlsson map with respect to the given reference tetrahedron.

To be more precise, it should be observed that five points (not four) define a projective coordinate frame in $\mathcal{P}^3$. In fact, there is more than one (in fact a 3-parameter family) of projective frames for which four non-coplanar points have coordinates $\mathbf{E}_i$. Thus the Carlsson map with respect to a given reference tetrahedron is not unique. However, the mapping given by definition 1 with respect to any such coordinate frame may be used.

Given a statement or theorem concerning projections of sets of points with respect to one or more projection centres one may derive a dual statement. One requires that among the four points being projected, there are four non-coplanar

points that may form a reference tetrahedron. Under a general duality mapping with respect to the reference tetrahedron

1. Points (other than those belonging to the reference tetrahedron) are mapped to centres of projection.
2. Centres of projection are mapped to points.
3. Straight lines are mapped to twisted cubics.
4. Ruled quadrics containing the reference tetrahedron are mapped to ruled quadrics containing the reference tetrahedron.

Points lying on an edge of the reference tetrahedron should be avoided, since the Carlsson mapping is undefined for such points. Using this as a sort of translation table, one may use existing theorems about point projection to be dualized, giving new theorems for which a separate proof is not needed.

Note : It is important to observe that only those points not belonging to the reference tetrahedron are mapped to camera centres by duality. The vertices of the reference tetrahedron remain points. In practice, in applying the duality principle, one may select any 4 points to form the reference tetrahedron, as long as they are non-coplanar. In general, in the results stated in the next section there will be an assumption (not always stated explicity) that **point sets considered contain four non-coplanar points**, which may be taken as the reference tetrahedron.

2.1 Reconstruction ambiguity

It will be shown in this section how various ambiguous reconstruction results may be derived simply from known, or obvious geometrical statements by applying duality.

We will be considering configurations of camera centres and 3D points, which will be denoted by $\{\mathbf{C}_1, \ldots, \mathbf{C}_m; \mathbf{X}_1, \ldots, \mathbf{X}_n\}$ or variations thereof. Implicit is that the symbols appearing before the semicolon are camera centres, and those that come after are 3D points. In order to make the statements of derived results simple, the concept of *image equivalence* is defined.

Definition 2. *Two configurations*

$$\{\mathbf{C}_1, \ldots \mathbf{C}_m; \mathbf{X}_1, \ldots \mathbf{X}_n\} \ and \ \{\mathbf{C}'_1, \ldots \mathbf{C}'_m; \mathbf{X}'_1, \ldots \mathbf{X}'_n\}$$

are called image equivalent *if for all i the image of the set of points* $\mathbf{X}_1, \ldots, \mathbf{X}_n$ *observed from camera centre* $\mathbf{C}_i$ *is projectively equivalent to the image of points* $\mathbf{X}'_1, \ldots, \mathbf{X}'_n$ *observed from* $\mathbf{C}'_i$.

This definition makes sense, only because an image is determined up to projective equivalence by the centre of projection. The image of the points $\mathbf{X}_1, \ldots, \mathbf{X}_n$ with respect to centre $\mathbf{C}_i$ may be thought of somewhat abstractly as the projective equivalence class of the set of rays $\{\overline{\mathbf{C}_i \mathbf{X}_j} : j = 1, \ldots, n\}$.

The concept of image equivalence is distinct from projective equivalence of the sets of points and camera centres involved. Indeed, the relevance of this to reconstruction ambiguity is that if a configuration $\{\mathbf{C}_1, \ldots, \mathbf{C}_m; \mathbf{X}_1, \ldots, \mathbf{X}_n\}$ allows another image-equivalent set which is not projective-equivalent, then this amounts to an ambiguity of the projective reconstruction problem, since the projective structure of the points and cameras is not uniquely defined by the set of images. In this case, we say that the configuration $\{\mathbf{C}_1, \ldots, \mathbf{C}_m, \mathbf{X}_1, \ldots \mathbf{X}_n\}$ *allows an alternative reconstruction.*

Single view ambiguity

As a simple example of what can be deduced using Carlsson duality, consider the following simple question : when do two points project to the same point in an image. The answer is obviously, when the two points lie on the same ray (straight line) through the camera centre. From this simple observation, one may deduce the following result.

(2.1.2). *Consider a set of camera centres* $\mathbf{C}_1, \ldots, \mathbf{C}_m$ *and a point* $\mathbf{X}_0$ *all lying on a single straight line. and let* $\mathbf{E}_i : i = 1, \ldots, 4$ *be the vertices of a reference tetrahedron. Let* $\mathbf{X}$ *be another point. The the two configurations*

$$\{\mathbf{C}_1, \ldots, \mathbf{C}_m; \mathbf{E}_1, \ldots, \mathbf{E}_4, \mathbf{X}\} \ and \ \{\mathbf{C}_1, \ldots, \mathbf{C}_m; \mathbf{E}_1, \ldots, \mathbf{E}_4, \mathbf{X}_0\}$$

are image-equivalent configurations if and only if $\mathbf{X}$ *lies on the same straight line.*

This is illustrated in Fig 1.

In passing to the dual statement, according to Theorem (2.0.1) the straight line becomes a twisted cubic through the four vertices of the reference tetrahedron. Thus the dual statement to ((2.1.2)) is :

(2.1.3). *Consider a set of points* $\mathbf{X}_i$ *and a camera centre* $\mathbf{C}_0$ *all lying on a single twisted cubic also passing through four reference vertices* $\mathbf{E}_k$. *Let* $\mathbf{C}$ *be any other camera centre. Then the configurations*

$$\{\mathbf{C}; \mathbf{E}_1, \ldots, \mathbf{E}_4, \mathbf{X}_1, \ldots, \mathbf{X}_m\} \ and \ \{\mathbf{C}_0; \mathbf{E}_1, \ldots, \mathbf{E}_4, \mathbf{X}_1, \ldots, \mathbf{X}_m\}$$

are image equivalent if and only if $\mathbf{C}$ *lies on the same twisted cubic.*

Since the points $\mathbf{E}_i$ may be any four non-coplanar points, and a twisted cubic can not contain 4 coplanar points, one may state this last result in the following form :

Proposition 1. *Let* $\mathbf{X}_1, \ldots, \mathbf{X}_m$ *be a set of points and* $\mathbf{C}_0$ *a camera centre all lying on a twisted cubic. Then for any other camera centre* $\mathbf{C}$ *the configurations*

$$\{\mathbf{C}; \mathbf{X}_1, \ldots, \mathbf{X}_m\} \ and \ \{\mathbf{C}_0; \mathbf{X}_1, \ldots, \mathbf{X}_m\}$$

are image equivalent if and only if $\mathbf{C}$ *lies on the same twisted cubic.*

This is illustrated in Fig **??**. It shows that camera pose can not be uniquely determined whenever all points and a camera centre lie on a twisted cubic.

Using similar methods one can show that this is one of only two possible ambiguous situations. The other case in which ambiguity occurs is when all points and the two camera centres lie in the union of a plane and a line. This arises as the dual of the case when the straight line through the camera centres meets one of the vertices of the reference tetrahedron. In this case, the dual of this line is also a straight line through the same reference vertex (see Theorem (2.0.1)), and all points must lie on this line or the opposite face of the reference tetrahedron. These results were brought to the attention of the computer-vision community by Buchanan ([**?**]).
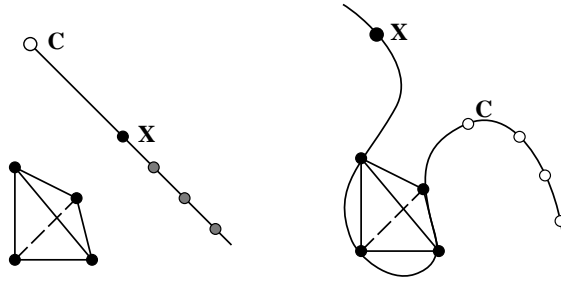

Fig. 1. Left : Any point on the line passing through **C** and **X** is projected to the same point from projection centre **C**.
Right : The dual statement – from any centre of projection **C** lying on a twisted cubic passing through **X** and the vertices of the reference tetrahedron, the five points are projected in the same way (up to projective equivalence). Thus a camera is constrained to lie on a twisted cubic by its image of five known points.

The horopter

Similar arguments can be used to derive the form of the horopter for two images. The horopter is the set of space points that map to the same point in two images. The following result is self-evident.

(2.1.4). *Given points* $\mathbf{X}$ *and* $\mathbf{X}'$*, the set of camera centres* $\mathbf{C}$ *such that*

$$\{\mathbf{C}; \mathbf{E}_1, \ldots, \mathbf{E}_4, \mathbf{X}\} \text{ and } \{\mathbf{C}; \mathbf{E}_1, \ldots, \mathbf{E}_4, \mathbf{X}'\}$$

are image equivalent is the straight line passing through $\mathbf{X}$ *and* $\mathbf{X}'$*.*

This is illustrated in Fig 2. The dual of this statement is

Proposition 2. *Given projection centres* $\mathbf{C}$ *and* $\mathbf{C}'$*, non-collinear with the four points* $\mathbf{E}_i$ *of a reference tetrahedron, the set of points* $\mathbf{X}$ *such that* $\{\mathbf{C}; \mathbf{E}_1, \ldots, \mathbf{E}_4, \mathbf{X}\}$ *and* $\{\mathbf{C}'; \mathbf{E}_1, \ldots, \mathbf{E}_4, \mathbf{X}\}$ *are image-equivalent is a twisted cubic passing through* $\mathbf{E}_1, \ldots, \mathbf{E}_4$ *and the two projection centres* $\mathbf{C}$ *and* $\mathbf{C}'$*.*
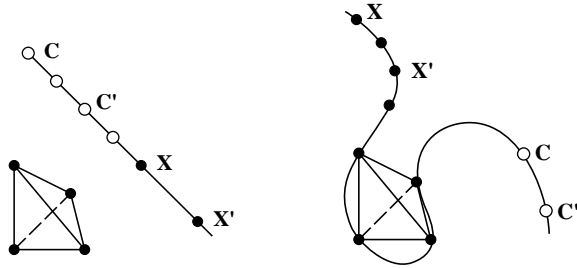
Fig. 2. Left : From any centre of projection $\mathbf{C}, \mathbf{C}', \ldots$ lying on the line passing through $\mathbf{X}$ and $\mathbf{X}'$, the points $\mathbf{X}$ and $\mathbf{X}'$ are projected to the same ray. That is, $\{\mathbf{C}; \mathbf{E}_i, \mathbf{X}\}$ is image-equivalent to $\{\mathbf{C}; \mathbf{E}_i, \mathbf{X}'\}$ for all $\mathbf{C}$ on the line.
Right : The dual statement – all points on the twisted cubic passing through $\mathbf{C}$ and $\mathbf{C}'$ and the vertices of the reference tetrahedron are projected in the same way relative to the two projection centres. That is, $\{\mathbf{C}; \mathbf{E}_i, \mathbf{X}\}$ is image-equivalent to $\{\mathbf{C}'; \mathbf{E}_i, \mathbf{X}\}$ for all $\mathbf{X}$ on the twisted cubic. This curve is called the *horopter* for the two centres of projection.

Note in both these examples how the use of duality has taken intuitively obvious statements concerning projections of points collinear points and derived a result somewhat less obvious about points lying on a twisted cubic.

Two-view ambiguity

The basic (well known) result about critical surfaces from two views may be stated as follows.

Theorem (2.1.5). *A configuration* $\{\mathbf{C}_1, \mathbf{C}_2; \mathbf{X}_1, \ldots, \mathbf{X}_n\}$ *of two camera centres and* n *points allows an alternative reconstruction if and only if both camera centres* $\mathbf{C}_1, \mathbf{C}_2$ *and all the points* $\mathbf{X}_j$ *lie on a ruled quadric surface. Furthermore, when an alternative reconstruction exists, then there will always exist a third distinct reconstruction.*

One may write down the dual statement straight away as follows.

Theorem (2.1.6). *A configuration* $\{\mathbf{C}_1, \ldots, \mathbf{C}_n; \mathbf{X}_1, \ldots, \mathbf{X}_6\}$ *of any number of cameras and six points allows an alternative reconstruction if and only if all camera centres* $\mathbf{C}_1, \ldots, \mathbf{C}_n$ *and all the points* $\mathbf{X}_1, \mathbf{X}_6$ *lie on a ruled quadric surface. Furthermore, when an alternative reconstruction exists, then there will always exist a third distinct reconstruction.*

This result was proven in [**?**]. Observe that in this dual statement, the value of n is not the same as the value of n in Theorem (2.1.5). Indeed, in the transition to the dual result, four of the original n points $\mathbf{X}_j$ are selected as the reference tetrahedron, and remain points. The remaining $n-4$ points become camera centres. The two original camera centres become points, making six points in total. The ruled quadric becomes a ruled quadric according to Theorem **??**.

9

The minimum interesting case of Theorem (2.1.6) is when $n = 3$, as studied in [?]. In this case one has nine points in total (three cameras and six points). One can construct a quadric surface passing through these nine points (a quadric is defined by nine points). If the quadric is a ruled quadric (a hyperboloid of one sheet in the non-degenerate case), then there are three possible distinct reconstructions. Otherwise the reconstruction is unique.

3 Dual Algorithms

The method of duality will now be given for deriving a dual algorithm from a given algorithm. Specifically, it will be shown that if one has an algorithm for doing projective reconstruction from n views of $m + 4$ points, then there is an algorithm for doing projective reconstruction from m views of $n + 4$ points. This result, observed by Carlsson [?], will be made specific by explicitly describing the steps of the dual algorithm.

We consider a projective reconstruction problem, which will be referred to as $\mathcal{P}(m, n)$. It is the problem of doing reconstruction from m views of n points. We denote image points by $\mathbf{x}_j^i$, which represents the image of the j-th object space point in the i-th view. Thus, the upper index indicates the view number, and the lower index represents the point number. Such a set of points $\{\mathbf{x}_j^i\}$ is called **realizable** if there are a set of camera matrices $\mathtt{P}^i$ and a set of 3D points $\mathbf{X}_j$ such that $\mathbf{x}_j^i = \mathtt{P}^i \mathbf{X}_j$. The projective reconstruction problem $\mathcal{P}(m, n)$ is that of finding such camera matrices $\mathtt{P}^i$ and points $\mathbf{X}_j$ given a realizable set $\{\mathbf{x}_j^i\}$ for m views of n points.

Let $\mathcal{A}(n, m + 4)$ represent an algorithm for solving the projective reconstruction problem $\mathcal{P}(n, m+4)$. An algorithm will now be exhibited for solving the projective reconstruction $\mathcal{P}(m, n + 4)$. This algorithm will be denoted $\mathcal{A}^*(m, n + 4)$, the dual of the algorithm $\mathcal{A}(n, m + 4)$.

Initially, the steps of the algorithm will be given without proof. In addition, difficulties will be glossed over so as to give the general idea without getting bogged down in details. In the description of this algorithm it is important to keep track of the range of the indices, and whether they index the cameras or the points. Thus, the following may help to keep track.

- Upper indices represent the view number.
- Lower indices represent the point number.
- i ranges from 1 to m.
- j ranges from 1 to n.
- k ranges from 1 to 4.

The dual algorithm

Given an algorithm $\mathcal{A}(n, m+4)$ the goal is to exhibit a dual algorithm $\mathcal{A}^*(m, n+ 4)$.

Input:

The input to the algorithm $\mathcal{A}^*(m, n+4)$ consists of a realizable set of $n+4$ points seen in m views. This set of points can be arranged in a table as in Fig 3 (left).
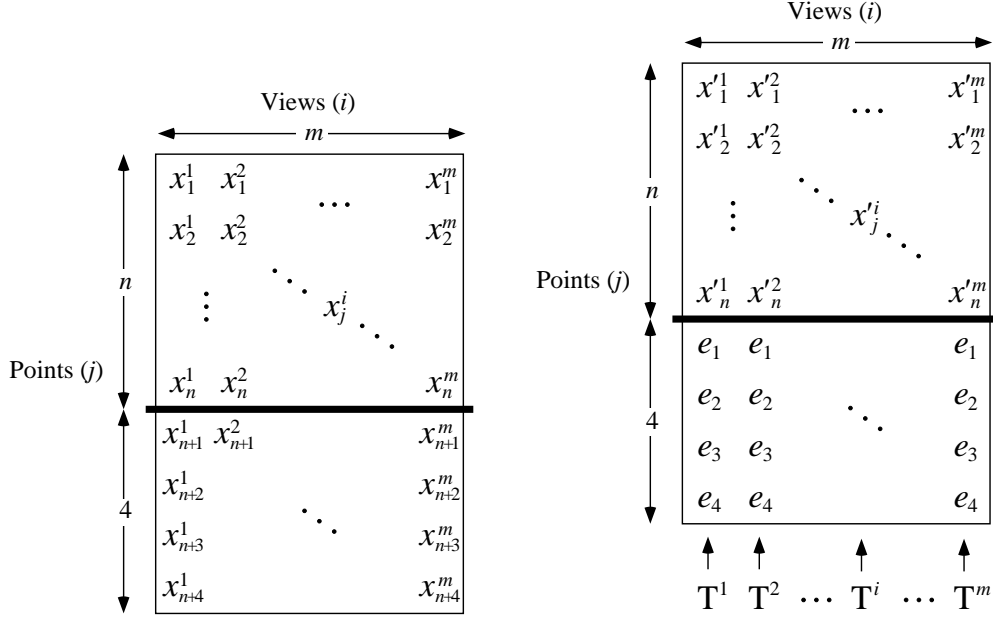

Fig. 3. Left : Input to algorithm $\mathcal{A}^*(m, n+4)$ **Right :** Input data after transformation.

In this table, the points $\mathbf{x}^i_{n+k}$ are separated from the other points $\mathbf{x}^i_j$, since they will receive special treatment.

Step 1 : Transform.

The first step is to compute for each i, a transformation $\mathtt{T}^i$ that map the points $\mathbf{x}^i_{n+k}$, $k = 1, \ldots, 4$ to the points $\mathbf{e}_k$ of a canonical basis for projective 2-space $\mathcal{P}^2$. The transformation $\mathtt{T}^i$ is applied also to each of the points $\mathbf{x}^i_j$ to produce transformed points $\mathbf{x}'^i_j = \mathtt{T}^i \mathbf{x}^i_j$. The result is the transformed point array shown in Fig 3(right). A different transformation $\mathtt{T}^i$ is computed and applied to each column of the array, as indicated.

Step 2 : Transpose.

The last four rows of the array are dropped, and the remaining block of the array is transposed. One defines $\hat{\mathbf{x}}^j_i = \mathbf{x}'^i_j$. At the same time, one does a mental

11

switch of points and views. Thus the point $\hat{\mathbf{x}}_i^j$ is now conceived as being the image of the j-th point in the i-th view, whereas the point $\mathbf{x}_j'^i$ was the image of the i-th point in the j-th view. What is happening here effectively is that one is swapping the roles of points and cameras – the basic concept behind Carlsson duality expressed by (3). The resulting transposed array is shown in Fig 4(left).
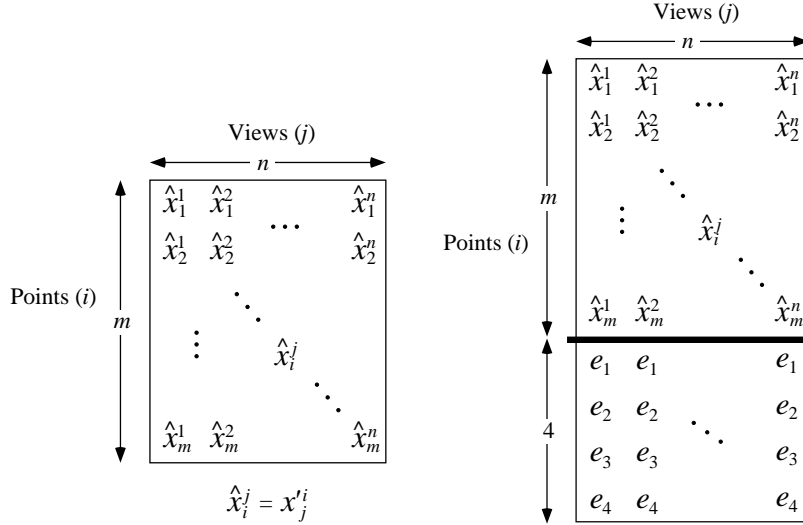

Fig. 4. Left : *Transposed data.* **Left :** Transposed data extended by addition of extra points.

Step 3 : Extend.

The array of points is now extended by the addition of four extra rows containing points $\mathbf{e}_k$ in all positions of the $(m + k)$-th row of the array, as shown in Fig 4(right).

Step 4 : Solve.

The array of points resulting from the last step has $m + 4$ rows and n columns, and may be regarded as the positions of $m + 4$ points seen in n views. As such, it is a candidate for solution by the algorithm $\mathcal{A}(n, m+4)$, which we have assumed is given. Essential here is that the points in the array form a realizable set of point correspondences. Justification of this is deferred for now. The result of the algorithm $\mathcal{A}(n, m + 4)$ is a set of cameras $\widehat{\mathsf{P}}^j$ and points $\widehat{\mathbf{X}}_i$ such that $\hat{\mathbf{x}}_i^j = \widehat{\mathsf{P}}^j \widehat{\mathbf{X}}_i$. In addition, corresponding to the last four rows of the array, there are points $\widehat{\mathbf{X}}_{m+k}$ such that $\mathbf{e}_k = \widehat{\mathsf{P}}^j \widehat{\mathbf{X}}_{m+k}$ for all j.

Step 5 : 3D transform.

Since the reconstruction obtained in the last step is a projective reconstruction, one may transform it (equivalently, choose a projective coordinate frame) such that the points $\widehat{\mathbf{X}}_{m+k}$ are the four points $\mathbf{E}_k$ of a partial canonical basis for $\mathcal{P}^3$. The only requirement is that the points $\widehat{\mathbf{X}}_{m+k}$ obtained in the projective reconstruction not be coplanar. This assumption is validated later.

At this point, one sees that $\mathbf{e}_k = \widehat{\mathsf{P}}^j \widehat{\mathbf{X}}_{m+k} = \widehat{\mathsf{P}}^j \mathbf{E}_k$. From this it follows that $\widehat{\mathsf{P}}^j$ has the special form

$$\widehat{\mathsf{P}}^j = \begin{bmatrix} a^j & & & d^j \\ & b^j & & d^j \\ & & c^j & d^j \end{bmatrix} \quad . \tag{4}$$

Step 6 : Dualize.

Let $\widehat{\mathbf{X}}_i = (\mathrm{X}_i, \mathrm{Y}_i, \mathrm{Z}_i, \mathrm{T}_i)^\top$, and $\widehat{\mathsf{P}}^j$ be as given in (4). Now define points $\mathbf{X}_j = (a^j, b^j, c^j, d^j)^\top$ and cameras

$$\mathsf{P}'^i = \begin{bmatrix} \mathrm{X}_i & & & \mathrm{T}_i \\ & \mathrm{Y}_i & & \mathrm{T}_i \\ & & \mathrm{Z}_i & \mathrm{T}_i \end{bmatrix} \quad .$$

Then one verifies that

$$\begin{aligned}
\mathsf{P}'^i \mathbf{X}_j &= (\mathrm{X}_i a^j + \mathrm{T}_i d^j, \mathrm{Y}_i b^j + \mathrm{T}_i d^j, \mathrm{Z}_i c^j + \mathrm{T}_i d^j)^\top \\
&= \widehat{\mathsf{P}}^j \widehat{\mathbf{X}}_i \\
&= \hat{\mathbf{x}}_i^j \\
&= \mathbf{x}_j'^i \quad .
\end{aligned}$$

If in addition, one defines $\mathbf{X}_{n+k} = \mathbf{E}_k$ for $k = 1, \ldots, 4$, then $\mathsf{P}'^i \mathbf{X}_{n+k} = \mathbf{e}_k$. It is then evident that the cameras P'^i and points $\mathbf{X}_j$ and $\mathbf{X}_{n+k}$ form a projective reconstruction of the transformed data array obtained in Step 1 of this algorithm.

Step 7 : Reverse transform.

Finally, defining $\mathsf{P}^i = (\mathrm{T}^i)^{-1} \mathsf{P}'^i$, and with the points $\mathbf{X}_j$ and $\mathbf{X}_{n+k}$ obtained in the previous step, one has a projective reconstruction of the original data. Indeed, one verifies

$$\mathsf{P}^i \mathbf{X}_j = (\mathrm{T}^i)^{-1} \mathsf{P}'^i \mathbf{X}_j = (\mathrm{T}^i)^{-1} \mathbf{x}_j'^i = \mathbf{x}_j^i \quad .$$

This completes the description of the algorithm. One can see that it takes place in various stages.

1. In Step 1, the data is transformed into canonical image reference frames based on the selection of 4 distinguished points.

2. In Steps 2 and 3 the problem is mapped into the dual domain, resulting in a dual problem $\mathcal{P}(n, m + 4)$.
3. The dual problem is solved in step 4 and 5.
4. Step 6 maps the solution back into the original domain.
5. Step 7 undoes the effects of the initial transformation.

3.1   Justification of the algorithm.

To justify this algorithm, one needs to be sure that at Step 4 there indeed exists a solution to the transformed problem. Before considering this, it is necessary to explain the purpose of Step 3, which extends the data by the addition of rows of image points $\mathbf{e}_k$, and Step 5, which transforms the arbitrary projective solution to one in which four points are equal to the 3D basis points $\mathbf{E}_k$.

The purpose of these steps is to ensure that one obtains a solution to the dual reconstruction problem in which $\widehat{\mathsf{P}}^j$ has the special form given by (4) in which the camera matrix is parametrized by only 4 values. The dual algorithm is described in this manner so that it will work with any algorithm $\mathcal{A}(n, m + 4)$ whatever. However, both Steps 3 and 5 may be eliminated if the known algorithm $\mathcal{A}(n, m+4)$ has the capability of enforcing this constraint on the camera matrices directly. Algorithms based on the fundamental matrix, trifocal or quadrifocal tensors may easily be modified in this way, as will be seen.

In the mean time, since $\widehat{\mathsf{P}}^j$ of the form (4) is called a *reduced camera matrix*, we call any reconstruction in which each camera matrix is of this form a *reduced reconstruction*. Not all sets of realizable point correspondences allow a reduced reconstruction, however, the following result characterizes sets of point correspondences that do have this property.

(3.1.7). *A set of image points $\{\mathbf{x}_j^i \ : \ i = 1, \ldots, m \ ; \ j = 1, \ldots, n\}$ permits a reduced reconstruction if and only if it may be augmented with supplementary correspondences $\mathbf{x}_{n+k}^i = \mathbf{e}_k$ for $k = 1, \ldots, 4$ such that*

1. *The total set of image correspondences is realizable, and*
2. *The reconstructed points $\mathbf{X}_{n+k}$ corresponding to the supplementary image correspondences are non-coplanar.*

Proof. The proof is straight-forward enough. Suppose the set permits a reduced reconstruction, and let P^i be the set of reduced camera matrices. Let points $\mathbf{X}_{n+k} = \mathbf{E}_k$ for $k = 1, \ldots, 4$ be projected into the m images. The projections are $\mathbf{x}_{n+k}^i = \mathsf{P}^i \mathbf{X}_{n+k} = \mathsf{P}^i \mathbf{E}_k = \mathbf{e}_k$ for all i.

Conversely, suppose the augmented set of points are realizable and the points $\mathbf{X}_{n+k}$ are non-coplanar. In this case, a projective basis may be chosen such that $\mathbf{X}_{n+k} = \mathbf{E}_k$. Then for each view, one has $\mathbf{e}_k = \mathsf{P}^i \mathbf{E}_k$ for all k. From this it follows that each P^i has the desired form (4). $\qquad\Box$

One other remark must be made before proving the correctness of the algorithm.

14

(3.1.8). *If a set of image points $\{\mathbf{x}_j^i \; : \; i = 1, \ldots, m \; ; \; j = 1, \ldots, n\}$ permits a reduced reconstruction then so does the transposed set $\{\hat{\mathbf{x}}_i^j \; : \; j = 1, \ldots, n \; ; \; i = 1, \ldots, m\}$ where $\hat{\mathbf{x}}_i^j = \mathbf{x}_j^i$ for all i and j.*

This is the basic duality property, effectively proven by the construction given in Step 6 of the algorithm above. Now it is possible to prove the correctness of the algorithm.

Proposition 3. *Let $\mathbf{x}_j^i$ and $\mathbf{x}_{n+k}^i$ as in Fig ?? be a set of realizable image point correspondences, and suppose*

1. *for each i, the four points $\mathbf{x}_{n+k}^i$ are non-collinear, and*
2. *the four points $\mathbf{X}_{n+k}$ in a projective reconstruction are non-coplanar.*

Then the algorithm of section ?? will succeed.

Proof. Because of the first condition of the theorem, transformations $\mathtt{T}^i$ exist for each i, transforming the input data to the form shown in Fig ??. This transformed data is also realizable, since the transformed data differ only by a projective transformation of the image.

Now, according to ((3.1.7)) applied to Fig ??, the correspondences $\mathbf{x}_j'^i$ admit a reduced realization. By ((3.1.8)) the transposed data Fig ?? also admits a reduced realization. Applying ((3.1.7)) once more shows that the extended data Fig ?? is realizable, Furthermore, the points $\widehat{\mathbf{X}}_{m+k}$ are non-coplanar, and so Step 5 is valid. The subsequent steps 6 and 7 go forward without problems. $\quad\square$

4 Refinements to the dual algorithm

The dual algorithm as presented above gives a way of dualizing any given projective reconstruction algorithm. The main weakness of this approach is that it ignores possible noise in the measurements. Noise ought to be considered at several points.

Direct enforcement of reduced reconstruction.

Steps 3 and 5 of the algorithm are used to make sure that the camera matrices in the computed reconstruction are of the form (4). The trouble with this is that the points $\hat{\mathbf{x}}_{m+k}^j = \mathbf{e}_k$ are treated as any other point in the reconstruction. In the presence of noise, most algorithms, such as those based on multifocal tensors find reconstructions for which the input point correspondences are only approximately satisfied, to the extent that is possible given the level of noise. However, in order that the camera matrices should be of the correct form, it is necessary that the correspondences $\hat{\mathbf{x}}_{m+k}^j = \mathbf{e}_k$ be satisfied exactly. Thus, these correspondences must be treated differently from the others.

Preferable would be to enforce the constraint that the camera matrices are of the form (4) directly. In the case where $n = 2$ the algorithm $\mathcal{A}(n, m + 4)$ used

to obtain the reconstruction in the dual domain may be the 8-point algorithm. Apart from assuming that each of the camera matrices is reduced, one may assume further that the first one has the special canonical form $\widehat{P}^1 = [I \mid 0]$. In this case with $\widehat{P}^2$ given as in (4) one computes that the fundamental matrix has the form (up to a scale factor)

$$\widehat{F} = \widehat{P}^2 = \begin{bmatrix} 0 & -b & c \\ a & 0 & -c \\ -a & b & 0 \end{bmatrix} \tag{5}$$

The 8-point algorithm may easily be modified so that the computed fundamental matrix has this form. The retrieval of the reduced camera matrix $\widehat{P}^2$ from (5) is then trivial.

In the case where $n = 3$, one may use an algorithm based on the trifocal tensor. For three general camera matrices $[I \mid 0]$, $A = [a_i^j]$ and $B = [b_i^k]$ the general formula for the trifocal tensor was given in [?] to be

$$T_i^{jk} = a_i^j b_4^k - a_4^j b_i^k \tag{6}$$

for $1 \leq i, j, k \leq 3$. Translated into the notation of the present paper and applied to reduced camera matrices $\widehat{P}^1 = [I \mid 0]$, $\widehat{P}^2$ and $\widehat{P}^3$ of the form (4) (and assuming that $d^1 = d^2 = 1$) one sees that there are only 15 non-zero entries of T_i^{jk} and these entries of T_i^{jk} are linear in terms of the values a^i, b^i and c^i for $i = 2, 3$. Thus, one may solve linearly for the T_i^{jk} corresponding to reduced camera matrices, and in fact find the entries of the reduced camera matrices linearly.

The transformations T^i

The most serious difficulty is finding a well-performing algorithm using this dualization scheme to reduce to a known algorithm is how to handle the transformations T^i. Application of projective transformations to the image data has the effect of distorting any noise distribution that may apply to the data. The problem also exists of choosing four points that are non-collinear in any of the images. If the points are close to collinear in any of the images, then the projective transformation applied to the image in Step 1 of the algorithm may entail extreme distortion of the image. In the algorithm discussed in [?] for computing the quadrifocal tensor, this sort of distortion was shown to degrade performance of the algorithm severely.

5 Experimental performance

Algorithms based on the fundamental matrix (the 8-point algorithm) for two views and the trifocal tensor (three views) were dualized, resulting in algorithms for 6 or 7 points in any number of views. The results of these tests were reported

as a student report in August 1996 by Gilles Debunne. Since this report is effectively unavailable, the results are summarized here.

Performance of the algorithms was generally unsatisfactory, mainly due to the distortion of the noise by the application of the transformations $\mathbf{T}^i$. It was observed that errors due to noise may be minimized in Step 4 of the algorithm. Reversing the dualization in Step 6 of the algorithm results in the same small errors. However, when the inverse projective transformations are applied in Step 7, the average error became very large. Some points retained quite small error, whereas in those images where distortion was significant, quite large errors resulted.

Normalization in the sense of [?] is also a problem. It has been shown to be essential for performance of the linear reconstruction algorithms to apply data normalization. However what sort of normalization should be applied to the transformed data of Fig ?? which is geometrically unrelated to actual image measurements is a mystery.

To get good results, it would seem that one would need to propagate assumed error distributions forward in Step 1 of the algorithm to get assumed error distributions for the transformed data Fig ??, and then during reconstruction to minimize residual error relative to this propagated error distribution. However, the fundamental matrix and trifocal tensor algorithms do not provide ways of dealing with arbitrary error distributions.

6   Conclusion

Duality as introduced by Carlsson is a very interesting theoretical tool for understanding camera projection. It seems also to have potential to provide algorithms for reconstruction from image sequences containing a large number of images. To this point, however, problems with dealing with noise distributions are an impediment to good performance.

There seems to be good hope, however for eventually using methods like this for finding linear algorithms for carrying out reconstruction from extended image sequences. Finding such a method would represent a significant advance, since at present linear methods for reconstruction have been limited to reconstruction from small numbers of views.