CENTER FOR
MACHINE PERCEPTION

CZECH TECHNICAL
UNIVERSITY

REPRINT

Oriented Matching Constraint

Tomáš Werner[1] and Tomáš Pajdla

werner@robots.ox.ac.uk, pajdla@cmp.felk.cvut.cz

[1] Visual Geometry Group, Department of Engineering Science, University of Oxford. Oxford, OX1 3PJ, UK. http://www.robots.ox.ac.uk/ vgg/

T.Werner and T.Pajdla. Oriented Matching Constraint. *In Proc. of the 12th British Machine Vision Conference, Manchester, UK. September 10-13, 2001.*

Center for Machine Perception, Department of Cybernetics
Faculty of Electrical Engineering, Czech Technical University
Technická 2, 166 27 Prague 6, Czech Republic
fax +420 2 2435 7385, phone +420 2 2435 7637, www: http://cmp.felk.cvut.cz

Oriented Matching Constraints

Tomáš Werner
Visual Geometry Group
University of Oxford
Oxford, OX1 3PJ, UK
http://www.robots.ox.ac.uk/˜vgg/

Tomáš Pajdla
Center for Machine Perception
Czech Technical University
Prague, 12135, Czech Republic
http://cmp.felk.cvut.cz

Abstract

Well-known matching constraints for points and lines in muliple images are necessary but not sufficient condition for the existence of real structure and cameras, underlying the image correspondences. To obtain sufficient conditions, the following additional constraints must be imposed: positive scales, the existence of a plane at infinity not intersecting the scene, and the existence of handedness preserving cameras. We present modifications of the well-known matching constraints and also some new constraints, taking into account some of this additional knowledge. Not only conventional but also central panoramic cameras are naturally described. To achieve this, we have generalized and simplified Hartley's ch(e)irality theory by formulating it in the language of oriented projective geometry and Grassmann tensors.

1 Introduction

In[1] multiple view geometry for 3D computer vision [2], projective space P_{d-1} and projective geometry are used as a theoretical background for the theory and algorithms. It has been shown [6, 4] that *oriented projective* space T_{d-1} (called spherical in mathematics) and the corresponding geometry are more natural to model the real Euclidean world than a projective one. In T_{d-1}, concepts like orientation of lines and planes, convex hull of points, etc., are well defined. Multiple view geometry can be easily augmented to live in oriented projective rather than projective space as follows: two homogeneous vectors, matrices or tensors representing points, lines, planes, conics, cameras, multifocal tensors, etc. are considered to represent the same geometrical object iff they are equal up to a positive rather than non-zero scale. Thus, the vectors $\mathbf{x}$ and $-\mathbf{x}$ represent different points.

The contribution of this paper is to augment multiple view matching constraints [2] with orientation. Using oriented projective geometry enables the consistent use of additional knowledge, available implicitly in almost every 3D computer vision task. Given matched correspondences in images, this knowledge is expressed by the following constraints:

Orientation consistency. Overall signs of vectors, matrices or tensors representing geometrical objects must not be changed and all scales must be positive. This says that we are in an oriented projective space rather than in a projective one.

[1] We acknowledge support of the European Commision FP5 (Marie Curie Fellowship MCFI-2000-01685, 1999-29017 OMNIVIEWS), the Czech Ministry of Education (MSMT KONTAKT 2001/09), and the Grant Agency of Czech Republic (102/01/0971). We also thank Bill Triggs for the discussion which has helped clarifying many ideas present explicitly or implicitly in this paper.

Scene affinity. There exists a plane (plane at infinity) in the scene such that all scene points are in front of it. This says that the true structure underlying the image correspondences is affine rather than oriented projective or projective.

Cameras preserve handedness. Very often the relation of handedness of image coordinate systems and the handedness of the true scene coordinate system is known (it can be either equal or opposite). The cameras are required to preserve this relation.

Directionality of conventional cameras. For conventional cameras, there is a line (line at infinity) in each image such that all image points are on its positive side. However, no such line exists for (central) panoramic cameras.

The paper builds on three works. The first is Stolfi's theory of oriented projected geometry of flats in arbitrary dimension [6]. Its usefulness for computer vision community has been noticed by Laveau and Faugeras [4]. However, extension to multiple view geometry is not straightforward since some concepts (e.g., of the camera) are missing.

The second is the excellent treatment of geometry of projective reconstruction and matching constraints due to Triggs [8]. Algebra of Grassmann tensors is used as the tool to represent the geometry of flats and projections. In fact, this algebra represents exactly the abstract algebra Stolfi uses as the language of oriented projective geometry: flats are represented by Grassmann tensors, and operations join and meet by antisymmetrizing over respectively contravariant and covariant indices. The orientation is distinguished by an overall sign of a tensor. However, geometrical insights due to considering orientation are not discussed explicitly in the paper, even if the author is aware of them. Personal communication with Bill Triggs [1] has helped us to clarify some important ideas.

The third is Hartley's theory of ch(e)irality and quasi-affine reconstruction [3, 2]. To our opinion, the existence of quasi-affine reconstruction is the most important result due to including the above constraints. However, the reasoning is done in projective geometry rather than, more naturally, in oriented projective one. An independent less exhaustive work on this topic is [10].

The paper is organized as follows. First, concepts and notation are introduced and their relation to [6, 8, 3] stated. Then, a hierarchy of possible reconstructions from image correspondences is presented, as the result of imposing the first, second and third of the above constraints. The hierarchy is closely related to weak and strong realization and quasi-affine reconstruction in [3, 2] but formulation in oriented projective space makes it simpler and the ideas are expressed in pure geometrical terms, without refering to any special coordinate system. The main part of the paper describes imposing the orientation consistency and partially also scene affinity and camera handedness preservation to well-known matching constraints in terms of multiple view tensors or joint image matrix [2]. Also entirely new matching constraints, as on oriented lines in two images, are presented.

2 Notation and concepts

As the paper is very closely related to the works [6, 8, 3], being familiar with them (especially with the first two) is *essential to understand the paper*. We use the same concepts and symbols as in [6, 8], without explaining them here again for the lack of space.

The geometrical entities involved in our considerations are flats of oriented projective spaces T_2 or T_3. The operations we need are as follows: taking antipode, join, meet, projective mapping, and generalized projective mapping. These are introduced and explained in Stolfi's book [6].

To represent these entities and operations, we use Grassmann algebra in terms of antisymmetric tensors. It is explained in [8] and we also use the same notation. This algebra naturally represents oriented projective geometry as follows: (flat of rank r in T_{d-1}, taking antipode, join, meet, projective mapping from T_{d-1} to T_{d-1}, generalized projective mapping from T_{d-1} to T_{e-1} for $e < d$) are represented respectively by (d-dimensional r-index contravariant Grassmann tensor, $-$, antisymmetrizing over contravariant indices, antisymmetrizing over covariant indices of dual tensor, transforming contravariant indices by $d \times d$ full rank matrix, transforming contravariant indices by $e \times d$ full rank matrix).

Any geometric object like a scene or image flat, camera matrix, multiple view tensor, etc., is oriented. All tensors differing just by an overall *positive* scale represent the same geometric object, while multiplying it by a negative scale yields the antipodal object. Equality up to a non-zero and positive scale is denoted respectively by $\simeq$ and $\doteqdot$.

Scene space is E_3 Euclidean 3-space but, since just affine rather than also metric properties of the scene are of interest in the paper, we consider it as A_3 affine 3-space. A_3 is isomorphic to the set of points of T_3 that are in front of a certain plane (plane at infinity) Ω [6]. This set is denoted by ordered pair $(T_3, \Omega) = A_3$ and represented by $(\mathcal{H}^a, \mathbf{\Omega}_a) = \{\mathbf{x}^a \in \mathcal{H}^a | \, \mathbf{\Omega}_a \mathbf{x}^a > 0\}$, where $\dim \mathcal{H}^a = 4$.

Scene is a set of points from (T_3, Ω) and flats of T_3 of rank higher than 1. It is represented by the set $\{\mathbf{x}_n^a, \mathbf{x}_m^{ab}, \mathbf{x}_l^{abc} | \, \mathbf{\Omega}_a \mathbf{x}_n^a > 0\}$ where the symbols $\mathbf{x}_n^a, \mathbf{x}_m^{ab}$, and $\mathbf{x}_l^{abc}$ denote respectively the n-th scene point, m-th scene line, and l-th scene plane. Since it is not possible to tell relative orientation (i.e., whether it is in front of or behind) of a flat of rank higher than 1 w.r.t. Ω (such flat always intersects Ω), there is no constraint like $\mathbf{\Omega}_a \mathbf{x}_n^a > 0$ for $\mathbf{x}_m^{ab}$ or $\mathbf{x}_l^{abc}$. Note, n, m, l are not meant as tensor indices but rather just to distinguish different flats, in other words, there is no vector space like $\mathcal{H}^n$.

By camera, we mean *central linear* camera, represented by a linear mapping in homogeneous coordinates. We distinguish *panoramic* and *directional* cameras and images. The field of view of a panoramic camera is $360°$ at least in one plane containing its center [5]. Directional camera sees only scene points in front of a certain plane Π (planar retina) through its center. It has the front range and inobservable back range [4]. There is no such plane for a panoramic camera, its retina is topologically a sphere. Conventional (e.g. TV or photographic) cameras are directional. Note, not every camera is central (i.e., its rays do not intersect in a single scene point) [5] and not every central camera is linear (i.e., the scene-to-image mapping is not linear in homogeneous coordinates) [9].

Image space is T_2 oriented projective 2-space, being in fact the set of directions in E_3 or the surface of a sphere in E_3. It is represented by $\mathcal{H}^A$ of dimension 3. Multiple image spaces are distinguished by index k and denoted as T_2^k and $\mathcal{H}^{A_k}$. *Image space of a directional camera* is a restriction of T_2 to affine 2-space $A_2 = (T_2, \pi)$, represented by the pair $(\mathcal{H}^A, \mathbf{\Pi}_A) = \{\mathbf{x}^A \in \mathcal{H}^A | \, \mathbf{\Pi}_A \mathbf{x}^A > 0\}$. $\mathbf{\Pi}_A$ is the image line at infinity. It is a projection of the planar retina Π, represented by $\mathbf{\Pi}_a = \mathbf{P}_a^A \mathbf{\Pi}_A$.

Image is a set of flats from T_2. It is represented by the set $\{\mathbf{x}_n^A, \mathbf{x}_m^{AB}, \mathbf{x}_l^{ABC}\}$ where the symbols denote respectively the n-th image point, m-th image line, and l-th image plane. Note, in T_2 (unlike in P_2) it is meaningful to consider also image planes because there are two oppositely oriented image planes in an image. *Directional image* is the set $\{\mathbf{x}_n^A, \mathbf{x}_m^{AB}, \mathbf{x}_l^{ABC} | \, \mathbf{\Pi}_A \mathbf{x}^A > 0\}$. Like in scene space, there is no constraint on relative orientation of image lines and planes w.r.t. $\mathbf{\Pi}_A$.

Camera is a generalized projective mapping [6] from scene space to image space. It is represented by a full-rank 3×4 matrix $\mathbf{P}_a^A$. *Directional camera* is a pair $(\mathbf{P}_a^A, \mathbf{\Pi}_A)$.

Camera center $\mathbf{e}^a$ is the null space of the camera mapping with uniquely chosen orientation, $\mathbf{e}^a = \frac{1}{3!}\varepsilon^{abcd}\mathbf{P}_b^A\mathbf{P}_c^B\mathbf{P}_d^C\varepsilon_{ABC}$ [8]. Projection of the center $\mathbf{e}_k^a$ of camera $\mathbf{P}_a^{A_k}$ by the camera $\mathbf{P}_a^{A_l}$ is *epipole* $\mathbf{e}_k^{A_l} = \mathbf{P}_a^{A_l}\mathbf{e}_k^a$.

We assume in the paper that the true cameras underlying the image correspondences *preserve handedness*. This can be assumed without loss of generality because if some of the true cameras reverses handedness, it suffices to mirror its image coordinate system prior to computations. The camera's handedness preserving property is a product of several other properties: (i) the way the camera solves visibility ('from the center' or 'towards the center'), (ii) whether images are direct or mirrored, and (iii) signatures of images and scene epsilon tensors [1]. Detailed explanation is rather subtle and omitted due to the lack of space.

Preserving handedness in fact means that the camera 'sees the positive side of Ω', $\mathbf{P}_a^A\mathbf{P}_b^B\mathbf{P}_c^C\mathbf{\Omega}^{abc} \doteq \varepsilon^{ABC}$. This in turn means that its center lies in front Ω, $\mathbf{\Omega}_a\mathbf{e}^a > 0$. Note, the orientation of the null space $\mathbf{e}^a$ of $\mathbf{P}_a^A$ is chosen so that for any plane $\mathbf{x}^{abc}$ it is $(\mathbf{P}_a^A\mathbf{P}_b^B\mathbf{P}_c^C\mathbf{x}^{abc} \doteq \varepsilon^{ABC}) \Leftrightarrow (\mathbf{x}_a\mathbf{e}^a > 0)$.

3 Hierarchy of reconstructions

Let $\{\mathbf{x}_n^{A_k}, \mathbf{x}_m^{A_kB_k}, \mathbf{x}_l^{A_kB_kC_k}\}$ be projections of N points, M lines and L planes in K images ($1 \le n \le N, 1 \le m \le M, 1 \le l \le L, 1 \le k \le K$). These image correspondences can originate from conventional or panoramic cameras, they are assumed to be correctly oriented[2], yet known only up to a positive scales. The correspondences are matched, thus e.g. $\mathbf{x}_1^{A_1}$ and $\mathbf{x}_1^{A_2}$ are projections of a single scene point in image 1 and 2. It is assumed that a true reconstruction (i.e., cameras and structure) exists, projecting to the correspondences.

We add several more steps into the well-known [2] hierarchy projective–affine–metric reconstruction. First, we define

Definition 1 *Projective reconstruction from the set of images* $\{\mathbf{x}_n^{A_k}, \mathbf{x}_m^{A_kB_k}, \mathbf{x}_l^{A_kB_kC_k}\}$ *is a set* $\{\mathbf{P}_a^{A_k}, \mathbf{x}_n^a, \mathbf{x}_m^{ab}, \mathbf{x}_l^{abc}\}$ *such that* $\mathbf{x}_n^{A_k} \simeq \mathbf{P}_a^{A_k}\mathbf{x}_n^a$, $\mathbf{x}_n^{A_kB_k} \simeq \mathbf{P}_a^{A_k}\mathbf{P}_b^{B_k}\mathbf{x}_n^{ab}$, *and* $\mathbf{x}_n^{A_kB_kC_k} \simeq \mathbf{P}_a^{A_k}\mathbf{P}_b^{B_k}\mathbf{P}_c^{C_k}\mathbf{x}_n^{abc}$.

Oriented projective reconstruction from $\{\mathbf{x}_n^{A_k}, \mathbf{x}_m^{A_kB_k}, \mathbf{x}_l^{A_kB_kC_k}\}$ *is defined similarly except that the symbols* $\simeq$ *are replaced by* $\doteq$.

Although stronger than projective reconstruction, oriented projective reconstruction says nothing about scene affinity or handedness preserving cameras. These constraints can be imposed on it, separately or simultaneously, thus obtaining more steps in the hierarchy of reconstructions. This is described in Section 5.

The strongest reconstruction that can be obtained knowing only that the true scene is affine and the true cameras preserve handedness is described by

[2]Practically, the orientations can be measured in images as follows. For *points in conventional cameras*, the orientations of all points is given by the fact that they all lie in front of the image line at infinity, $\mathbf{\Pi}_a\mathbf{x}_n^A > 0$. If $\mathbf{\Pi}_a = (0,0,1)$ as usual, it means that the last coordinates of $\mathbf{x}_n^A$ are positive. In *panoramic cameras*, an image point and its antipode are clearly distinguished by their very design – e.g., for a camera consisting of a curved mirror and a conventional camera, a point and its antipode are two different points in the image of the mirror in the conventional camera. Orientations of *lines* can be obtained during the line detection algorithm from the direction of the intensity gradient orthogonal to the line. Even if this direction can swap when lighting conditions change severely, it can be expected highly invariant.

Theorem 1 *True scene, cameras and plane at infinity are in the set of all reconstructions* $\{\mathbf{P}_a^{A_k}, \mathbf{x}_n^a, \mathbf{x}_m^{ab}, \mathbf{x}_l^{abc}, \mathbf{\Omega}_a\}$ *satisfying: (i)* $\{\mathbf{P}_a^{A_k}, \mathbf{x}_n^a, \mathbf{x}_m^{ab}, \mathbf{x}_l^{abc}\}$ *is an oriented projective reconstruction from* $\{\mathbf{x}_n^{A_k}, \mathbf{x}_m^{A_k B_k}, \mathbf{x}_l^{A_k B_k C_k}\}$*, (ii)* $\mathbf{\Omega}_a \mathbf{x}_n^a > 0$*, and (iii)* $\mathbf{\Omega}_a \mathbf{e}_k^a > 0$*.*

4 Imposing oriented projective reconstructibility

Well-known matching constraints in terms of multilinear relations are (up to singularities) equivalent only to the existence of a projective reconstruction. This section shows how the condition on the existence of an oriented projective reconstruction can be imposed on matching constraints on flats in multiple images.

4.1 Points in two images

Let $\mathbf{x}^a$ be a scene point and $\mathbf{P}_a^{A_1}, \mathbf{P}_a^{A_2}$ two cameras. It is easy to verify (also Table 2 in [8]) that $\mathbf{e}_1^{A_2} \mathbf{x}^{B_2} \varepsilon_{A_2 B_2 C_2} \bumpeq \mathbf{x}^{C_1} \mathbf{F}_{C_1 C_2}$ (in familiar matrix notation [2], $\mathbf{e}' \times \mathbf{x}'_n \bumpeq \mathbf{F}\mathbf{x}_n$), where $\mathbf{F}_{C_1 C_2} = \varepsilon_{A_1 B_1 C_1} \mathbf{P}_a^{A_1} \mathbf{P}_b^{B_1} \varepsilon^{abcd} \mathbf{P}_c^{A_2} \mathbf{P}_d^{B_2} \varepsilon_{A_2 B_2 C_2}$ is the fundamental matrix and $\mathbf{x}^{A_k} = \mathbf{P}_a^{A_k} \mathbf{x}^a$. Hence (see Table 1 in the full version of [8], and also [9])

Theorem 2 *The image points* $\mathbf{x}_n^{A_k}$ *in two images (*$k = 1, 2$*) possess an oriented projective reconstruction if and only if there exists a matrix* $\mathbf{F}_{C_1 C_2}$ *of rank 2 and a non-zero vector* $\mathbf{e}_1^{A_2}$ *such that* $\mathbf{F}_{C_1 C_2} \mathbf{e}_1^{C_2} = \mathbf{0}_{C_1}$ *and for all* n *it is* $\mathbf{e}_1^{A_2} \mathbf{x}_n^{B_2} \varepsilon_{A_2 B_2 C_2} \bumpeq \mathbf{x}_n^{C_1} \mathbf{F}_{C_1 C_2}$*.*

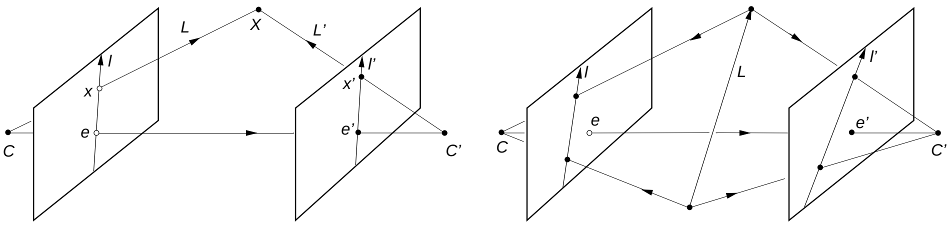


Figure 1: Constraint on points (left subfigure) and lines (right subfigure) in two images.

The geometric meaning of the theorem is shown in Fig. 1 on the left. The oriented ray denoted by L in the figure is given by $\mathbf{x}_{ab}^1 = \mathbf{x}_{A_1 B_1} \mathbf{P}_a^{A_1} \mathbf{P}_b^{B_1}$. However, L is also the join of the camera center C and the scene point X. L, C and X are projected into the second image as l', e' and x'. The condition requires that l' is the join of e' and x'.

4.2 Lines in two images

Let $\mathbf{x}^{ab}$ be a scene line and $\mathbf{P}_a^{A_1}, \mathbf{P}_a^{A_2}$ two cameras. It can be shown that $\mathbf{x}_{A_1} \mathbf{e}_2^{A_1} + \mathbf{x}_{A_2} \mathbf{e}_1^{A_2} = 0$ (in matrix notation [2], $\mathbf{l}\mathbf{e} + \mathbf{l}'\mathbf{e}' = 0$), where $\mathbf{x}_{C_k} = \mathbf{P}_a^{A_k} \mathbf{P}_b^{B_k} \mathbf{x}^{ab} \varepsilon^{A_k B_k C_k}$ are projections of the scene line. We have proven (the proof omitted for the limited space)

Theorem 3 *The set* $\{\mathbf{x}_{A_1}^n, \mathbf{x}_{A_2}^n\}$ *of* N *oriented image lines in two images possesses an oriented projective reconstruction if and only if there exist arbitrary non-zero 3-vectors* $\mathbf{e}_2^{A_1}$ *and* $\mathbf{e}_1^{A_2}$ *such that for every* n *it is* $\operatorname{sgn}(\mathbf{x}_{A_1}^n \mathbf{e}_2^{A_1}) + \operatorname{sgn}(\mathbf{x}_{A_2}^n \mathbf{e}_1^{A_2}) = 0$*. If this is satisfied for some* $\mathbf{e}_2^{A_1}$ *and* $\mathbf{e}_1^{A_2}$*, it is also satisfied for* $\mathbf{e}_2^{A_1}$ *and* $\mathbf{e}_1^{A_2}$ *equal to the epipoles of the image pair.*

The geometrical interpretation of Theorem 3 is shown in the right subfigure of Fig. 1. Relative orientation of the scene line L w.r.t. the line CC' is observed in each image as relative orientation of the image line l resp. l' w.r.t. the epipole e resp. e'. The constraint requires consistency of these two relative orientations in images.

The theorem can be formulated also as follows. Let us allow arbitrary pair-wise changes of image line orientations, meaning in fact changing the orientations of the underlying scene lines. *A set of image line pairs has consistent orientations if and only if the line orientations can be pair-wise reversed such that there exists a point in each image lying on a single side of each line.* In Fig. 2, two pair-wise changes have been done, achieving the situation when the epipole is on a single side of each line in each image. Note in the final image pair, the image lines rotate around each epipole, meaning that the scene lines merely rotate around the line connecting the camera centers.

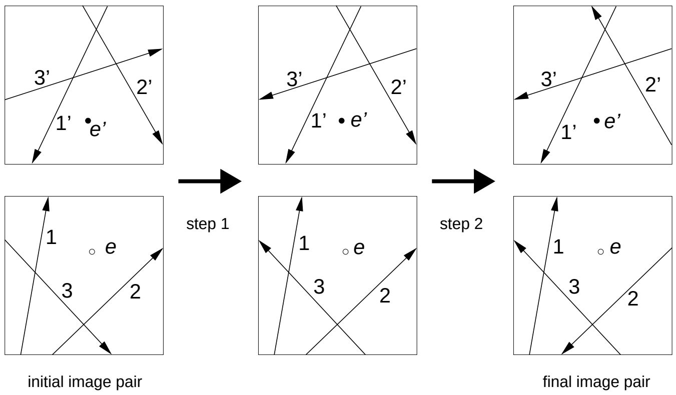


Figure 2: Sequence of pair-wise changes of line orientations, finished by the situation in which e is on a single (right-hand) side of each line in the first image, and e' is on a single (left-hand) side of each line in the second image.

If the epipoles are known, testing for the condition in Theorem 3 is easy. Not given the epipoles, it might seem that the condition could be tested by linear programming. But it is not so, we search simultaneously for epipoles and for reversions of orientations of underlying scene lines. We have not found any polynomial algorithm yet. If it were available, the epipoles could be approximately located from oriented lines in two images.

If the epipoles are known (up to orientations), the minimum number of line pairs not satisfying the condition in Theorem 3 is obviously two. What is the minimal number of lines not satisfying this condition for *unknown epipoles*?

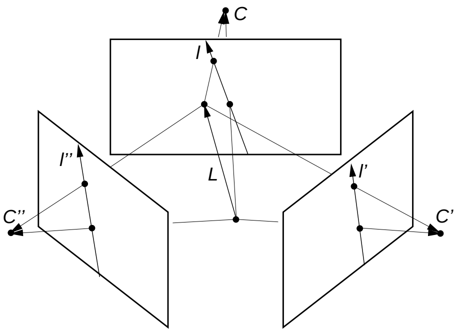
4.3 Lines in three images

The well-known bilinear relation for line transfer from two images to the third one, $\mathbf{x}_{A_3} = \mathbf{T}_{A_3}^{A_1 A_2} \mathbf{x}_{A_1} \mathbf{x}_{A_2}$, cannot yield correct orientations of the line $\mathbf{x}_{A_3}$ for the following simple reason. If the signs $\mathbf{x}_{A_1}$ and $\mathbf{x}_{A_2}$ are changed simultaneously, $\mathbf{x}_{A_3}$ remains unchanged. However, this is incorrect because reversing orientation of the first two image lines means reversing the underlying scene line, so the third image line must reverse too. Clearly, the correct line transfer formula must be a relation more complicated than a bilinear form.

The orientation consistent formula for reconstruction of a scene line from two image lines can be shown to be $\mathbf{x}_{ab} = (\mathbf{e}_1^{B_2} \mathbf{x}_{B_2})^{-1} \mathbf{x}_{A_1} \mathbf{x}_{A_2} \mathbf{P}_{[a}^{A_1} \mathbf{P}_{b]}^{A_2}$. Since $\mathbf{e}_1^{B_2} \mathbf{x}_{B_2} + \mathbf{e}_2^{B_1} \mathbf{x}_{B_1} = 0$, the scalar in parentheses could be also $-\mathbf{e}_2^{B_1} \mathbf{x}_{B_1}$, the complete symmetry cannot be achieved. Ignoring singularities, we obtain the correct transfer equation by simply projecting the scene line $\mathbf{x}_{ab}$ to a third image, $\mathbf{x}_{A_3} = (\mathbf{e}_1^{B_2}\mathbf{x}_{B_2})^{-1}\mathbf{x}_{A_1}\mathbf{x}_{A_2}\mathbf{T}_{A_3}^{A_1A_2}$, where $\mathbf{T}_{A_3}^{A_1A_2} = \mathbf{P}_a^{A_1}\mathbf{P}_b^{A_2}\mathbf{P}_c^{B_3}\mathbf{P}_d^{C_3}\varepsilon^{abcd}\varepsilon_{A_3B_3C_3}$ is the trifocal tensor. Hence

Theorem 4 *For N corresponding lines $\mathbf{x}_{A_k}^n$ in three images it holds $\mathbf{e}_1^{B_2}\mathbf{x}_{B_2}^n\mathbf{x}_{A_3}^n \doteq \mathbf{x}_{A_1}^n\mathbf{x}_{A_2}^n\mathbf{T}_{A_3}^{A_1A_2}$, where $\mathbf{T}_{A_3}^{A_1A_2}$ is the trifocal tensor.*

The situation is shown in Fig. 3 on the left. Having changed orientation of the scene line L, all three image lines l, l', l'' must change their orientations too.



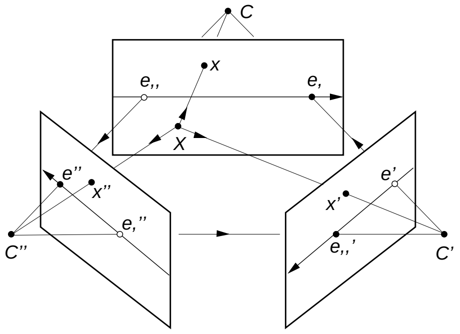


Figure 3: Constraint on lines (left subfigure) and points (right subfigure) in three images.

4.4 Points in three images

The well-known constraint $\mathbf{x}^{A_1}\varepsilon_{A_1B_1C_1}\mathbf{x}^{A_2}\varepsilon_{A_2B_2C_2}\mathbf{T}_{C_3}^{C_1C_2}\mathbf{x}^{C_3} = 0_{B_1B_2}$ [2], allowing point transfer via trifocal tensor, does not determine the orientation of $\mathbf{x}^{C_3}$. The orientation could be determined using pairs of images and Theorem 2. However, we present a different constraint, which is inherent to a triplet of images. Let $\mathbf{x}^a$ be a scene point. Then it can be shown that

$$\mathbf{x}^a\mathbf{e}_1^b\mathbf{e}_2^c\mathbf{e}_3^d\varepsilon_{abcd} = \mathbf{x}^{A_1}\mathbf{e}_2^{B_1}\mathbf{e}_3^{C_1}\varepsilon_{A_1B_1C_1} = \mathbf{x}^{A_2}\mathbf{e}_3^{B_2}\mathbf{e}_1^{C_2}\varepsilon_{A_2B_2C_2} = \mathbf{x}^{A_3}\mathbf{e}_1^{B_3}\mathbf{e}_2^{C_3}\varepsilon_{A_3B_3C_3}$$

or, in matrix language, $[\mathbf{X}, \mathbf{C}^1, \mathbf{C}^2, \mathbf{C}^3] = [\mathbf{x}^1, \mathbf{e}_2^1, \mathbf{e}_3^1] = [\mathbf{x}^2, \mathbf{e}_3^2, \mathbf{e}_1^2] = [\mathbf{x}^3, \mathbf{e}_1^3, \mathbf{e}_2^3]$, where $[\mathbf{x}, \mathbf{y}, \mathbf{z}]$ denotes the determinant of the matrix with columns $\mathbf{x}, \mathbf{y}, \mathbf{z}$. Hence

Theorem 5 *Let $\mathbf{x}_n^{A_k}$, $k = 1, 2, 3$, possess an oriented projective reconstruction. Then there are image lines $\mathbf{t}_{A_k}$ such that $\mathbf{t}_{A_k}\mathbf{e}_i^{A_k} = 0$ for $k, i = 1, 2, 3$, $i \neq k$, (i.e., the epipoles lie on the lines) and $\mathrm{sgn}(\mathbf{t}_{A_1}\mathbf{x}_n^{A_1}) = \mathrm{sgn}(\mathbf{t}_{A_2}\mathbf{x}_n^{A_2}) = \mathrm{sgn}(\mathbf{t}_{A_3}\mathbf{x}_n^{A_3})$ for all n.*

Geometrical interpretation is shown in Fig. 3 on the right. Relative orientation of the trifocal plane $CC'C''$, represented by $\mathbf{e}_1^a\mathbf{e}_2^b\mathbf{e}_3^c\varepsilon_{abcd}$, w.r.t. the scene point X is observed in each image as relative orientation of the line joining the two epipoles w.r.t. the image point.

4.5 Points in joint image matrix

Recall [8] that projection of a scene point $\mathbf{x}^a$ by K cameras can be viewed as projection by a big $3K \times N$ joint camera matrix $\mathbf{P}_a^\alpha$ to a single $3K$-dimensional image point $\mathbf{P}_a^\alpha\mathbf{x}^a$, where $\mathcal{H}^\alpha$ is $3K$-dimensional *joint image space*. For multiple points measured in the images with correct orientations it is $\rho_N^k\mathbf{x}_n^{A_k} = \mathbf{P}_a^k\mathbf{x}_n^a$, i.e., in matrix form,

$$\begin{pmatrix} \rho_1^1\mathbf{x}_1^{A_1} & \dots & \rho_N^1\mathbf{x}_N^{A_1} \\ \vdots & \dots & \vdots \\ \rho_1^K\mathbf{x}_1^{A_K} & \dots & \rho_N^K\mathbf{x}_N^{A_K} \end{pmatrix} = \begin{pmatrix} \mathbf{P}_a^{A_1} \\ \vdots \\ \mathbf{P}_a^{A_K} \end{pmatrix} \begin{pmatrix} \mathbf{x}_1^a & \dots & \mathbf{x}_N^a \end{pmatrix}.$$

Let the *joint image matrix* on the left be denoted by $(\rho_n^k \mathbf{x}_n^{A_k})$. This matrix must have rank at most 4, which is the principle of projective reconstruction by factorization [7]. Clearly,

Theorem 6 *Let $\mathbf{x}_n^{A_k}$ be N image points measured in K images with correct orientations. An oriented projective reconstruction from $\mathbf{x}_n^{A_k}$ exists if and only if there are $\rho_n^k > 0$ such that* $\operatorname{rank}(\rho_n^k \mathbf{x}_n^{A_k}) \leq 4$.

5 Imposing scene affinity and preserving handedness

Imposing scene affinity and handedness preserving cameras on matching constraints is more difficult than imposing positive scales. We present some interesting results for points in special situations.

Scene affinity can be imposed only on points (an oriented projective reconstruction with all points in front of Ω is called *quasi-affine* in [3, 2]) because one cannot say whether an infinite line or plane lies in front of Ω. A meaningful constraint would be obtained if they were considered as point sets rather than flats. However, requiring cameras to preserve handedness is meaningful even for a reconstruction not containing any points.

5.1 Using camera directionality

Camera directionality is an additional knowledge implying the Theorems 7 and 8. These theorems have been presented in [3, 2], we give here their more general and compact form. They can be used to strengthen Theorems 2, 5 and 6.

Theorem 7 *Let images $\mathbf{x}_n^{A_k}$ of N points in K cameras have an oriented projective reconstruction and let at least one camera be directional. Then there is an affine scene projecting to $\mathbf{x}_n^{A_k}$ by some (not necessarily handedness preserving) cameras.*

Proof. Knowing that $\mathbf{\Pi}_a^k \mathbf{x}^a > 0$, let $\mathbf{\Omega}_a$ be one of the image planes $\mathbf{\Pi}_a^k$ or their convex linear combination, which implies $\mathbf{\Omega}_a \mathbf{x}^a > 0$. ■

It follows that if points $\mathbf{x}_n^{A_k}$ in multiple images possess an oriented projective reconstruction and one or both cameras are directional, there is an affine scene projecting to $\mathbf{x}_n^{A_k}$. Thus, Theorems 2, 5 and 6 and directionality of one or more cameras imply the existence of an affine scene projecting to $\mathbf{x}_n^{A_k}$ by some (not necessarily handedness preserving) cameras.

Theorem 8 *Let images $\mathbf{x}_n^{A_k}$ of N points in 2 cameras have an oriented projective reconstruction and let at least one camera be directional. Then there is an affine scene projecting to $\mathbf{x}_n^{A_k}$ by handedness preserving cameras.*

Proof. Due to Theorem 7, there is an affine scene projecting to $\mathbf{x}_n^{A_k}$, i.e., there exist $\mathbf{\Omega}_a$ such that $\mathbf{\Omega}_a \mathbf{x}_n^a > 0$. The cameras must satisfy $\mathbf{\Omega}_a \mathbf{e}_k^a > 0$. In fact, condition $(\mathbf{\Omega}_a \mathbf{e}_1^a)(\mathbf{\Omega}_a \mathbf{e}_2^a) > 0$ is sufficient because the case $\mathbf{\Omega}_a \mathbf{e}_k^a < 0$ can be brought to $\mathbf{\Omega}_a \mathbf{e}_k^a > 0$ by letting $\mathbf{x}_n^a \mapsto \mathbf{x}_n^{a'} = \mathbf{H}_a^{a'} \mathbf{x}_n^a$ and $\mathbf{P}_a^A \mapsto \mathbf{P}_{a'}^A = \mathbf{P}_a^A (\mathbf{H}^{-1})_{a'}^a$ for some $\mathbf{H}_a^{a'}$ with negative determinant. If $(\mathbf{\Omega}_a \mathbf{P}_1^a)(\mathbf{\Omega}_a \mathbf{P}_2^a) < 0$ and at least one camera is directional, it is always possible to move the plane $\mathbf{\Omega}_a$ so that one of the camera centers becomes separated from the other center and the scene points, as illustrated in Fig. 4. ■

In other words, oriented projective reconstructibility and directionality of one or both cameras are necessary and sufficient for the existence of a real scene and real cameras underlying $\mathbf{x}_n^{A_k}$ [9].

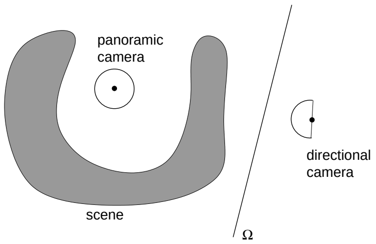


Figure 4: Having two cameras, at least one of them directional, the directional camera center can always be separated from the scene points and the second camera center by plane.

5.2 Imposing affinity on points in joint image matrix

Recall Section 4.5. If all cameras are panoramic, Theorem 6 does not guarantee the existence of an affine scene projecting to $\mathbf{x}_n^{A_k}$. We present a constraint sufficient for the existence of an affine scene (but not necessarily handedness preserving cameras).

The existence of an affine scene requires the existence of a plane at infinity $\mathbf{\Omega}_a$ such that $\mathbf{\Omega}_a\mathbf{x}^a > 0$. Let $\mathbf{\Omega}_\alpha$ be a hyperplane in the joint image space $\mathcal{H}^\alpha$ satisfying $\mathbf{\Omega}_\alpha\mathbf{x}^\alpha > 0$. Then its preimage in the scene, $\mathbf{\Omega}_a = \mathbf{\Omega}_\alpha\mathbf{P}_a^\alpha$, will satisfy $\mathbf{\Omega}_a\mathbf{x}^a > 0$. Hence

Theorem 9 *Let $\mathbf{x}_n^{A_k}$ (or equivalently $\mathbf{x}_n^\alpha$) possess an oriented projective reconstruction. An affine scene projecting to $\mathbf{x}_n^{A_k}$ exists if and only if there is $\mathbf{\Omega}_\alpha$ such that $\mathbf{\Omega}_\alpha\mathbf{x}^\alpha > 0$.*

This is an interesting result saying that the *joint camera is directional*, its image plane being the plane at infinity $\mathbf{\Omega}_a$.

6 Conclusion

Well-known matching constraints for points and lines are only *necessary* conditions for the existence of any real geometry projecting to the given image points or lines. Besides singularities, the reason is that they are projective rather than affine, even if the underlying geometry is indeed affine. Based on oriented projective geometry represented by Grassman tensors and the ch(e)irality theory, we have modified the old constraints and presented some entirely new constraints. The modified constraints are stronger than the original ones. Some are even *necessary and sufficient* for the existence of a real scene and cameras.

Of course, there is a general way how to test any set of corresponding image points for the existence of an underlying real geometry: compute a projective reconstruction and solve ch(e)iral inequalities [3, 2] by e.g. linear programming. However, our task was to obtain these constraints in terms of *image entities*, like image points, image lines, epipoles and multiview tensors. This has allowed achieving new *geometrical insights* in multiple view geometry. Moreover, some constraints are entirely new.

To be able to achieve results for matching constraints, we formulated the ch(e)irality theory [3, 2] in oriented projective geometry, generalized it for panoramic cameras and flats of arbitrary rank, and represented in Grassmann tensor algebra. We believe that this synthesis is a step towards a rigorous theory of oriented multiple view geometry. Such theory will become more necessary when panoramic cameras are used frequently [1], since some concepts and problems that are trivial for conventional cameras need formalization for more general cameras. Moreover, the theory applies quite straightforwardly for different dimensions of the scene and image spaces, e.g. for cameras represented by 2×4 projection matrix, which have a line as its 'projection center' and another line as a 'retina', and, in fact, for an arbitrary dimensions of scene and image spaces.

Using this new framework, we have not only derived new results (Theorems 3, 4, 5, 6, and 9) but also simplified and/or generalized some existing results (Definition 1 and Theorems 1, 7 and 8 from [3, 9], and Theorem 2 from [8, 9]).

The paper is not easy to read. Besides the fact that the topic itself is difficult, the reason is that the limited space of this paper allowed only for very compact text. To explain the topic in a more reader-friendly way would require much longer paper with many figures. We refer to [6] instead. However, the definitions and theorems should be independent on the (possibly too brief and unclear) rest of the text, as long as the reader is familiar with the notation from [8].

There is a number of open questions. Polynomial algorithm and minimal counter-example for lines in two images with unknown epipoles. Perhaps, Theorem 3 is a tautology: using a Monte Carlo search, we did not succeed in finding any configuration of image lines violating Theorem 3. Constraints on points and lines in more images can be combined. Is there a closed form of matching constraints involving panoramic cameras with scene affinity and camera handedness imposed? Even if the paper is meant as theoretical, quantifying the usefulness of the constraints for practical tasks would be interesting. Among these tasks are constraining search space in matching, discarding solutions for multiview tensors not corresponding to any real geometry (e.g., in estimating $\mathbf{F}$ or $\mathbf{T}_{A_3}^{A_1 A_2}$ from minimal number of correspondences), restricting the set of solutions for structure and cameras, and oriented flat transfer.

References

[1] E-mail communication with Bill Triggs, spring 2001.

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

[3] R. I. Hartley. Chirality. *Int. Jour. Computer Vision IJCV*, 26(1):41–61, 1998.

[4] S. Laveau and O. Faugeras. Oriented projective geometry for computer vision. In *ECCV96*, pages I:147–156, 1996.

[5] T. Pajdla, T. Svoboda, and V. Hlaváč. Epipolar geometry of central panoramic cameras. In R. Benosman and S. B. Kang, editors, *Panoramic Vision : Sensors, Theory, and Applications*. Springer Verlag, Berlin, Germany, 2000.

[6] J. Stolfi. *Oriented Projective Geometry: A Framework for Geometric Computations*. Academic Press, Inc., 1250 Sixth Avenue, San Diego, CA 92101, 1991.

[7] P. Sturm and B. Triggs. A factorization based algorithm for multi-image projective structure and motion. In *ECCV96*, pages II:709–720, 1996.

[8] B. Triggs. Matching constraints and the joint image. In *ICCV95*, pages 338–343, 1995. Also the full paper in the author's www home page.

[9] T. Werner and T. Pajdla. Cheirality in epipolar geometry. In *Proc. Intl. Conf. Computer Vision*. IEEE Computer Society Press, July 2001.

[10] T. Werner, T. Pajdla, and V. Hlaváč. Oriented projective reconstruction. In M. Gengler, M. Prinz, and E. Schuster, editors, *Pattern Recognition and Medical Computer Vision: 22-nd Workshop of the Austrian Association for Pattern Recognition (ÖAGM/IAPR)*, pages 245–254, Wien, Austria, May 14–15 1998. Österreichische Computer Gesselschaft.