View-synthesis from Uncalibrated Cameras and Parallel planes

Antonio Canclini[b], Francesco Malapelle[a], Marco Marcon[b], Stefano Tubaro[b], Andrea Fusiello[a,*]

[a] *DPIA - University of Udine, Italy*

[b] *DEIB - Politecnico di Milano, Italy*

Abstract

In this paper we present a view synthesis algorithm for the generation of novel intermediate views that smoothly interpolate between two reference images of the same scene acquired from two different point of views. If the acquisition system is not calibrated, there is no way to access the Euclidean frame and therefore it is impossible to determine the position and the orientation of the cameras as well as the three dimensional reconstruction of the imaged scene. Under the assumption of equal internal calibration of the two cameras, it is proved that the knowledge of the infinite homography is a sufficient condition for generating physically valid synthetic images from a pair of uncalibrated images. The view synthesis problem is therefore brought to the simpler well-known task of the estimation of the infinite homography through a set of geometric constraints existing in the scene. The work is complemented by the introduction of two novel techniques for estimating the infinite homography. This contributes to provide a variety of possibilities for such estimation, depending on the geometric constraints that can be identified in the imaged scene. The validity of the work is supported by a set of simulations on a synthetic three-dimensional scene and experiments on real data as well.

Keywords: View-synthesis; Image based rendering; Infinite homography;

1. Introduction

The term *View Synthesis* refers to the process of generating virtual images starting from a sparse set of reference images of a scene acquired by cameras in different positions. The novel images are called virtual views, as they are supposed to correspond to a picture of the scene taken from a virtual view point by a virtual camera. In the literature three classes of view synthesis methodologies can be identified. The first one relies on the concept of model based rendering. The algorithms belonging to this category are based on an explicit geometrical 3D model of the scene and therefore require the full calibration (internal and external parameters) of the acquisition system. Once the 3D model is reconstructed, the novel views are synthesized by specifying the virtual point of view and defining the extrinsic parameters (rotation and translation) of the virtual camera. The accuracy of the virtual view depends on the accuracy of the 3D model. Some examples can be found in [11, 14, 6, 21, 7]. The second class of the techniques for the view synthesis can be identified in the methods based on the plenoptic function [15, 20], which describes the intensity of the light observed from every position and direction in the 3D space. These methods do not require - in principle – any geometric information on the acquired objects, nor the computation of point correspondences. However, they need a large set of reference images. The third class of view synthesis techniques relies on uncalibrated reference views. When the acquisition system cannot be calibrated, i.e., when no a priori information on the camera positions and on their internal parameters is available, only a projective reconstruction of the 3D scene can be obtained. In other words, one can only reconstruct a 3D model affected by an unknown projective transformation, which is characterized by 15 degrees of freedom. The view synthesis process consists in this case in the interpolation among the set of reference views, by exploiting the geometric constraints existing in the images. The milestone of this approach is the work by Shashua and Navab ([3, 2]), which constitutes the basis of the further developments by Fusiello [8, 4] and Tebaldini [22].

All the uncalibrated view-synthesis algorithms require the specification of some projective elements, like epipoles, homographies, fundamental matrices or tensors. With the exception of the aforementioned papers, previous work concentrates on the generation of the novel view, assuming that the problem of view specification is solved somehow. For example, in [18, 3], the user has to manually specify the position of four points in each frame of the synthetic sequence. In [12] images are assumed to be nearly rectified, so the view synthesis reduces to the interpolation

*Corresponding author

Email addresses: name.surname@polimi.it (Antonio Canclini), name.surname@uniud.it (Francesco Malapelle), name.surname@polimi.it (Marco Marcon), name.surname@polimi.it (Stefano Tubaro), name.surname@uniud.it (Andrea Fusiello)

of horizontal disparities. The view specification problem is addressed in [5], where the internal parameters are assumed to be approximately known, thereby violating the assumption of uncalibrated camera.

The work presented in this paper deals with the view synthesis starting from two uncalibrated views of the scene. The focus on this class of view synthesis techniques reveals to be very promising for modern applications such as virtual navigation systems, where many pictures of the same scene are taken from different positions. These systems allow the user to virtually walk through the streets of a city, just connecting in a quite rough way the available set of pictures. The major limitation of these applications is that the images always rotate around a fixed observation point, and consequently it is impossible to switch between two view points in a continuous way. This problem could be solved by employing a view synthesis algorithm that continuously and smoothly interpolates between pairs of reference views, generating novel views as they would be taken from virtual cameras. In these kind of applications the view synthesis process cannot rely on a complete 3D representation of the scene, either because it is unpractical to obtain it, or because it is impossible to access the calibration information of the acquisition systems.

Starting from the concept of the relative affine structure introduced by Shashua and Navab [19], in this paper we revisit and extend the work by Tebaldini [22], deriving a complete framework for the generation of novel views that smoothly interpolates the reference views. This is made possible since the rigid motion of an object can be described by means of matrices that belong to a particular algebraical group, called special euclidean group ($SE(3)$). This is a Lie group and allows to calculate a smooth trajectory that interpolates between the initial and the final position of the camera. In principle, a virtual camera can be specified exponentiating the matrix that maps the transformation between the two views of the scene. However, this procedure can be directly applied only in the case of calibrated cameras, i.e. when all the metric information about the observed 3D scene is known. In the more general case of uncalibrated cameras, the approach has to be extended and adapted to the information contained in the images. We will see that the problem of the synthesis of virtual views can be definitively solved from the knowledge of the infinite homography matrix, which is the transformation induced by the plane at infinity. This is true under the assumption that the reference views are taken by cameras having equal internal calibration; however this does not constitute a big limitation, since in general multiple pictures are taken by means of a single camera without changing the focal properties (no zooming for example). The estimation of the infinite homography is well known in Computer Vision, and several techniques are proposed in the literature. In this paper we also propose two novel solutions to this issue, by exploiting some geometric constraints detectable in the two views, such as planes parallelism. Parallel planes are indeed common in most built environments, so this technique is relevant to virtual navigation systems such as Google Street View and Microsoft GeoSynth.

With respect to [22] in this paper we provide a more comprehensive formalization of the view synthesis problem and improve on some practical aspects. In particular: i) we prove that the knowledge of the homography induced by the plane at the infinity is sufficient for determining the synthetic views; ii) we get rid of reference 3D points that were needed in [22]. Moreover, we assessed the effectiveness of the method by means of extensive simulated experiments and in a real setup as well, by comparing the coordinates of the points of the synthetic views with those obtained from a ground-truth 3D model.

The rest of the paper is organized as follows. Section 2 gives an overview on the concepts of projective geometry, such as a review on the multi-view geometry and on the infinite homography, which constitute the background for understanding the paper. Section 3 is devoted to illustrate the notion of *relative affine structure*, which rest at the basis of the view synthesis algorithm. In particular the algorithm proposed by [19] will be reviewed, which requires to specify the position of the virtual camera in a projective frame. The solution to this problem relies on the concepts of Linear Lie Groups and Lie Algebra, which will be depicted in Section 4, with the purpose of generating smooth interpolating trajectories in the projective frame. Section 5 turns all the previous notions to account and describes the view synthesis methodology based on the knowledge of the infinite homography between two uncalibrated views which is the core of this paper. The section starts by showing how the knowledge of the infinite homography is a sufficient condition for the synthesis of new views. The section is completed by the description of the two novel methods for estimating the infinite homography from geometric elements visible in the two views, namely two parallel planes and a vanishing point or two pairs of parallel planes. Section 6 illustrates the experimental evaluation of our approach and at last in Section 7 conclusions are drawn.

2. Planar homography and the special case of infinite homography

This section reviews some basic concepts of multiple view geometry in Computer Vision. In particular we focus on the concept of planar homography, or equivalently, of homography induced by a plane, which constitutes the basis of the relative affine structure introduced in the next section. For a comprehensive dissertation the reader is referred to [16].

Consider a 3D point $\mathbf{X}$ lying on generic plane $\boldsymbol{\pi}$, as depicted in Figure 1. The planar homography is the projective transformation $\mathbf{A}_{ij}(\pi)$ that transfers the image of the point on the first view $\mathbf{x}_i$ to its projection on the sec-

ond view $\mathbf{x}_j$, i.e.[1]

$$\mathbf{x}_j \cong \mathbf{A}_{ij}(\pi)\mathbf{x}_i \ .$$

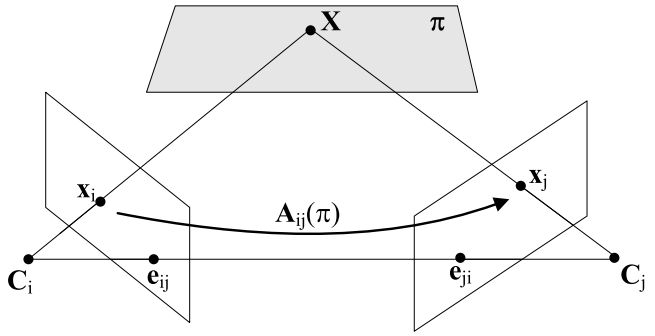


Figure 1: Planar homography: the image on the first view $\mathbf{x}_i$ of a 3D point $\mathbf{X}$ lying on the plane $\boldsymbol{\pi}$ is mapped to the point $\mathbf{x}_j$ in the second view through the transformation $\mathbf{A}_{ij}(\pi)$.

When the homography is induced by the plane at infinity $\boldsymbol{\pi}_\infty = [0,0,0,1]^T$, it takes the name of *infinite homography*. The view synthesis procedure described in Section 5 assumes the knowledge of this matrix, whose expression can be derived as follows. Consider a generic pair of camera matrices

$$\begin{aligned} \mathbf{P}_i &= \mathbf{K}_i[\mathbf{R}_i|\mathbf{t}_i] \\ \mathbf{P}_j &= \mathbf{K}_j[\mathbf{R}_j|\mathbf{t}_j] \ . \end{aligned}$$

Let $\mathbf{x}_i$ be a point on the first view. In the reference frame associated to the first camera, correspondent to the canonical camera matrix $\bar{\mathbf{P}}_i = \mathbf{K}_i[\mathbf{I}|\mathbf{0}]$, the ray passing through $\mathbf{x}_i$ is directed as the unit vector

$$\mathbf{d}_i = \frac{\mathbf{K}_i^{-1}\mathbf{x}_i}{||\mathbf{K}_i^{-1}\mathbf{x}_i||} \ .$$

Therefore, the vector $[\mathbf{d}_i, 0]^T$ represents the intersection between the ray and $\boldsymbol{\pi}_\infty$. In the reference frame associated to the second camera, considering $\bar{\mathbf{P}}_j = \mathbf{K}_j[\mathbf{I}|\mathbf{0}]$, the same direction is given by $\mathbf{d}_j = \mathbf{R}_{ij}\mathbf{d}_i$, with $\mathbf{R}_{ij} = \mathbf{R}_j\mathbf{R}_i^{-1}$ being the relative rotation between the two cameras. The point at infinity is then $[\mathbf{d}_j, 0]^T$, whose projection onto the second view is

$$\mathbf{x}_j \cong \mathbf{K}_j[\mathbf{I}|\mathbf{0}] \begin{bmatrix} \mathbf{d}_j \\ 0 \end{bmatrix} = \mathbf{K}_j\mathbf{d}_j \cong (\mathbf{K}_j\mathbf{R}_{ij}\mathbf{K}_i^{-1})\mathbf{x}_i \ . \quad (1)$$

From Equation (1) is easy to verify that the homography induced by the plane at infinity is given by

$$\mathbf{H}_{ij}^\infty \cong \mathbf{K}_j\mathbf{R}_{ij}\mathbf{K}_i^{-1} \ .$$

In the following we will always refer to the infinite homography with the matrix

$$\mathbf{A}_{ij}^\infty = \mathbf{K}_j\mathbf{R}_{ij}\mathbf{K}_i^{-1} \ ; \quad (2)$$

this scaling constitutes the base of the view synthesis algorithm described in Section 5. If the two calibration matrices are equal, i.e. $\mathbf{K}_i = \mathbf{K}_j$, then $\det(\mathbf{A}_{ij}^\infty) = 1$. Another important result is that an affine reconstruction can be calculated from the infinite homography [16]. More specifically, the reconstruction of the epipolar geometry obtained with the camera matrices $\mathbf{P}_i = [\mathbf{I}|\mathbf{0}]$ and $\mathbf{P}_j = [\mathbf{A}_{ij}^\infty|\mathbf{e}_{ji}]$ is affine.

3. Relative affine structure

In this section we show that the relation between the point correspondences in two images can be expressed by means of the homography induced by a generic reference plane, plus an affine term which multiplies the epipole of the second view. First of all we will show that this is true for the plane at infinity $\boldsymbol{\pi}_\infty$. Then, the result is extended for a generic plane $\boldsymbol{\pi}$.

Consider two cameras with generic projection matrices $\mathbf{P}_i = \mathbf{K}_i[\mathbf{R}_i|\mathbf{t}_i]$ and $\mathbf{P}_j = \mathbf{K}_j[\mathbf{R}_j|\mathbf{t}_j]$, whose centres are $\mathbf{C}_i = [\tilde{\mathbf{C}}_i, 1]^T$ and $\mathbf{C}_j = [\tilde{\mathbf{C}}_j, 1]^T$. Consider the point $\mathbf{X}$ and its projections $\mathbf{x}_i \cong \mathbf{P}_i\mathbf{X}$ and $\mathbf{x}_j \cong \mathbf{P}_j\mathbf{X}$ in the two views; the projections are scaled in such a way that $\mathbf{x}_i = [\tilde{\mathbf{x}_i}^T, 1]^T$ and $\mathbf{x}_j = [\tilde{\mathbf{x}_j}^T, 1]^T$. In the reference system of the first camera, for which the projection matrix is $\mathbf{P}_i^{(i)} = \mathbf{K}_i[\mathbf{I}|\mathbf{0}]$, the point $\mathbf{x}_i$ is obtained as

$$\mathbf{x}_i \cong z_i\mathbf{x}_i = \mathbf{P}_i^{(i)} \begin{bmatrix} z_i\mathbf{K}_i^{-1}\mathbf{x}_i \\ 1 \end{bmatrix} \ ,$$

where the superscript $^{(i)}$ is used to stress that we are working in the reference frame of the first camera. Due to the particular scaling of $\mathbf{x}_i$, $z_i\mathbf{K}_i^{-1}\mathbf{x}_i$ are the Euclidean coordinates of the point $\mathbf{X}$ in the reference system of the first camera. Moreover, z_i is the depth of $\mathbf{X}$, i.e. the signed distance of $\mathbf{X}$ from the principal plane. On the other hand, the reference system of second camera is characterized by the projection matrix $\mathbf{P}_j^{(j)} = \mathbf{K}_j[\mathbf{I}|\mathbf{0}]$. As before, the superscript $^{(j)}$ indicates that the reference system is associated to the second camera. It is easy to verify that

$$\begin{aligned} \mathbf{x}_j &\cong \mathbf{P}_j^{(j)} \begin{bmatrix} \mathbf{R}_{ij}(z_i\mathbf{K}_i^{-1}\mathbf{x}_i) + \mathbf{t} \\ 1 \end{bmatrix} \\ &= z_i\mathbf{A}_{ij}^\infty\mathbf{x}_i + \mathbf{K}_j\mathbf{t} \\ &\cong \mathbf{A}_{ij}^\infty\mathbf{x}_i + \frac{1}{z_i}\mathbf{K}_j\mathbf{t} \ , \end{aligned} \quad (3)$$

where $\mathbf{t} = \mathbf{t}_j - \mathbf{t}_i$ is the vector joining the camera centres expressed in Euclidean coordinates. Since in this reference system the camera centres are given by

$$\begin{aligned} \mathbf{C}_j^{(j)} &= [\mathbf{0}^T, 1]^T \\ \mathbf{C}_i^{(j)} &= \begin{bmatrix} \mathbf{R}_j & \mathbf{t}_j \\ \mathbf{0}^T & 1 \end{bmatrix} \mathbf{C}_i \ , \end{aligned}$$

it is readily verified that $\mathbf{t} = [\mathbf{R}_j|\mathbf{t}_j]\mathbf{C}_i$ and therefore

$$\mathbf{e}_{ji} \cong \bar{\mathbf{e}}_{ji} = \mathbf{K}_j\mathbf{t} = \mathbf{K}_j[\mathbf{R}_j|\mathbf{t}_j]\mathbf{C}_i = \mathbf{P}_j\mathbf{C}_i \ ,$$

[1] By default, all the points are in homogeneous coordinates. Cartesian coordinates will be denoted by a tilde above the variable.

where $\bar{\mathbf{e}}_{ji}$ refers to the specific scale[2] of the epipole $\mathbf{e}_{ji}$. Finally we obtain

$$\mathbf{x}_j \cong \mathbf{A}_{ij}^{\infty}\mathbf{x}_i + \mu\bar{\mathbf{e}}_{ji} \ , \tag{4}$$

with $\mu = \frac{1}{z_i}$. The term μ is called *relative affine structure*.

Let us now focus on the case in which a general plane $\boldsymbol{\pi} = [\mathbf{v}^T, c]^T$ is considered as a reference. This vector is homogeneous and therefore it can be scaled in such a way that $\boldsymbol{\pi} = [\mathbf{n}^T, -d_\pi]^T$, where $\mathbf{n} = \frac{v}{||v||}$ is the unit vector normal to the plane and $-d_\pi$ is the signed distance of the plane from the origin. Recalling the expression of the infinite homography given in Equation (2), from Equation (3) we have

$$\mathbf{x}_j \cong \mathbf{K}_j\mathbf{R}_{ij}\mathbf{K}_i^{-1}\mathbf{x}_i + \frac{\mathbf{K}_j\mathbf{t}}{z_i} \ . \tag{5}$$

Consider again the point $\mathbf{X}$ and its Euclidean representation in the reference system of the first camera, i.e. $z_i\mathbf{K}_i^{-1}\mathbf{x}_i$. We distinguish two cases:

- If $\mathbf{X}$ lies on the plane $\boldsymbol{\pi}$, since the centre $\mathbf{C}_i^{(i)}$ coincides with the origin, it follows that the distance of the plane from $\mathbf{C}_i^{(i)}$ is

  $$d_\pi = \mathbf{n}^T(z_i\mathbf{K}_i^{-1}\mathbf{x}_i)$$

  which is the projection on the plane normal $\mathbf{n}$ of the reprojected point $\mathbf{x}_i$. As a consequence, Equation (5) can be rewritten as

  $$\begin{aligned}\mathbf{x}_j &\cong \mathbf{K}_j\mathbf{R}_{ij}\mathbf{K}_i^{-1}\mathbf{x}_i + \frac{\mathbf{K}_j\mathbf{t}}{z_i}\frac{d_\pi}{d_\pi} \\ &= \mathbf{K}_j\mathbf{R}_{ij}\mathbf{K}_i^{-1}\mathbf{x}_i + \mathbf{K}_j\frac{\mathbf{t}\mathbf{n}^T}{d_\pi}\mathbf{K}_i^{-1}\mathbf{x}_i \\ &= \mathbf{K}_j\left(\mathbf{R}_{ij} + \frac{\mathbf{t}\mathbf{n}^T}{d_\pi}\right)\mathbf{K}_i^{-1}\mathbf{x}_i \ . \end{aligned} \tag{6}$$

  From Equation (6) we recognize the homography induced by the plane $\boldsymbol{\pi}$

  $$\mathbf{A}_{ij}(\pi) \cong \mathbf{K}_j\left(\mathbf{R}_{ij} + \frac{\mathbf{t}\mathbf{n}^T}{d_\pi}\right)\mathbf{K}_i^{-1} \ . \tag{7}$$

- Let $\mathbf{X}$ be a point not on the plane. Adding and subtracting the quantity $\mathbf{K}_j\frac{\mathbf{t}\mathbf{n}^T}{d_\pi}\mathbf{K}_i^{-1}\mathbf{x}_i$ to Equation (5) and rearranging the terms we obtain

  $$\mathbf{x}_j \cong \mathbf{A}_{ij}(\pi)\mathbf{x}_i + \left(\frac{d_\pi - \mathbf{n}^T(z_i\mathbf{K}_i^{-1}\mathbf{x}_i)}{d_\pi z_i}\bar{\mathbf{e}}_{ji}\right) \ .$$

  Since $d = d_\pi - \mathbf{n}^T z_i\mathbf{K}_i^{-1}\mathbf{x}_i$ is the signed distance between $\mathbf{X}$ and $\boldsymbol{\pi}$, it follows that

  $$\mathbf{x}_j \cong \mathbf{A}_{ij}(\pi)\mathbf{x}_i + \mu\bar{\mathbf{e}}_{ji} \tag{8}$$

  with

  $$\mu = \frac{d}{z_i d_\pi}$$

  as depicted in Figure 2

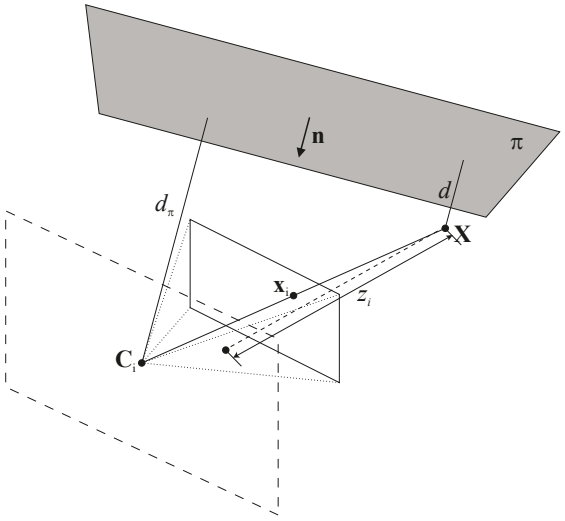


Figure 2: Representation of different terms in the relative affine structure for a point $\mathbf{X}$ out of the reference plane π.

  In order to solve the scaling ambiguity, Shashua and Navab [19] proposed to normalize both $\mathbf{A}_{ij}(\pi)$ and μ such that $\mu = 1$ for a certain finite point $\mathbf{X}^0$ not lying on the reference plane. This way, after the normalization one obtains the expression for the relative affine structure

  $$\mu = \frac{d}{z_i}\frac{z_i^0}{d^0} \ ,$$

  where d^0 is the distance of $\mathbf{X}^0$ from $\boldsymbol{\pi}$; and z_i^0 is the depth of $\mathbf{X}^0$ with respect to the first camera.

We summarize here some interesting properties of the relative affine structure:

- $\mu = 0$ for each point on the reference plane. For those points, in fact, $d_i = 0$ and $\mathbf{x}_j \cong \mathbf{A}_{ij}(\pi)\mathbf{x}_i$;
- if the plane at infinity is chosen as a reference, then $d, d_0, d_\pi \to \infty$; therefore $\frac{d}{d_0} \to 1$ and

  $$\begin{aligned}\mathbf{A}_{ij}(\pi) &\to \mathbf{A}_{ij}^{\infty} \\ \mu &\to \frac{z_i^0}{z_i} \ .\end{aligned}$$

- the relative affine structure depends only on the distance of $\mathbf{X}$ from the reference plane and on the position of the first camera, while it is independent on the second camera;
- for any choice of $\boldsymbol{\pi}$

  $$\mathbf{e}_{ji} \cong \mathbf{A}_{ij}(\pi)\mathbf{e}_{ij} + \mu\bar{\mathbf{e}}_{ji} \cong \mathbf{A}_{ij}(\pi)\mathbf{e}_{ij} \ ;$$

  since the epipoles are aligned with the cameras centers along the baseline.

[2] We recall that the epipole is represented by a homogeneous vector, i.e. $\mathbf{e}_{ji} \cong \lambda\mathbf{e}_{ji}\ \forall\lambda \neq 0$. In order to refer to a particular scaling of the epipole, we use the notation $\bar{\mathbf{e}}_{ji} = \mathbf{P}_j\mathbf{C}_i$. Notice that the quantity $\bar{\mathbf{e}}_{ji}$ can not be treated as a homogeneous vector.

- expanding Equation (7), we easily obtain a relation between the homography induced by $\boldsymbol{\pi}$ and the infinite homography:

$$\mathbf{A}_{ij}(\pi) \cong \mathbf{A}_{ij}^{\infty} + (\mathbf{K}_j\mathbf{t})\frac{\mathbf{n}^T\mathbf{K}_i^{-1}}{d_\pi} \ .$$

By expressing the plane $\boldsymbol{\pi}$ in the reference system of the first camera, i.e. $\boldsymbol{\pi}_i = [\mathbf{v}_i^T, c_i]^T \cong [\mathbf{n}^T\mathbf{K}_i^{-1}, -d_\pi]^T$, we also have that

$$\frac{\mathbf{v}_i^T}{c_i} = -\frac{\mathbf{n}^T\mathbf{K}_i^{-1}}{d_\pi} \ .$$

Since $\bar{\mathbf{e}}_{ji} = \mathbf{K}_j\mathbf{t}$, we finally obtain

$$\mathbf{A}_{ij}(\pi) \cong \mathbf{A}_{ij}^{\infty} - \bar{\mathbf{e}}_{ji}\frac{\mathbf{v}_i^T}{c_i} \ . \tag{9}$$

3.1. Relative affine structure and view synthesis

The model of the relative affine structure described by Equation (8) allows to build a 3D representation of the scene geometry by means of the pair of camera matrices in canonical form

$$\hat{\mathbf{P}}_i = [\mathbf{I}|\mathbf{0}] \ ; \quad \hat{\mathbf{P}}_j = [\mathbf{A}_{ij}(\pi)|\mathbf{e}_{ji}] \ .$$

The reconstruction is given by

$$\mathbf{X}^P = [\mathbf{x}_i^T, \mu]^T \ ,$$

and it is immediate to see that

$$\mathbf{x}_i \cong \hat{\mathbf{P}}_i\mathbf{X}^P \ ; \quad \mathbf{x}_j \cong \hat{\mathbf{P}}_j\mathbf{X}^P \ .$$

For the projective reconstruction ambiguity, for a general choice of the reference plane, the 3D reconstruction given by $\mathbf{X}^P$ is affected by an unknown projective transformation. As noticed in Section 2, a special case is given when the plane at infinite is selected as a reference. In this case the 3D reconstruction is affine.

Shashua and Navab proposed a methodology for the synthesis of new views based on the relative affine structure. The algorithm can be summarized as follows:

1. from a sufficiently populated set of point correspondences in the two views $(\mathbf{x}_i^{(k)}, \mathbf{x}_j^{(k)})$ estimate the fundamental matrix $\mathbf{F}$ and the epipoles by solving

$$\begin{aligned} \mathbf{x}_j^{(k)T}\mathbf{F}\mathbf{x}_i^{(k)} &= 0 \ , \forall k \\ \mathbf{F}^T\mathbf{e}_{ji} &= 0 \\ \mathbf{F}\mathbf{e}_{ij} &= 0 \ ; \end{aligned} \tag{10}$$

2. determine the reference plane $\boldsymbol{\pi}$ and the associated homography $\mathbf{A}_{ij}(\pi)$ by selecting (even randomly) three pairs of correspondences:

$$\begin{aligned} \mathbf{x}_j^{(k)} &\cong \mathbf{A}_{ij}(\pi)\mathbf{x}_i^{(k)} \ , k = 1, 2, 3 \\ \mathbf{e}_{ji} &\cong \mathbf{A}_{ij}(\pi)\mathbf{e}_{ij} \ ; \end{aligned}$$

3. calculate the relative affine structure

$$\mu_k = -\frac{\left([\mathbf{x}_j^{(k)}]_\times\mathbf{e}_{ji}\right)^T\left([\mathbf{x}_j^{(k)}]_\times\mathbf{A}_{ij}(\pi)\mathbf{x}_i^{(k)}\right)}{\left\|[\mathbf{x}_j^{(k)}]_\times\mathbf{e}_{ji}\right\|^2} \tag{11}$$

and reconstruct the scene geometry

$$\mathbf{X}_k^P \cong \begin{bmatrix}\mathbf{x}_i^{(k)} \\ \mu_k\end{bmatrix} \ ; \tag{12}$$

4. specify the matrix of the virtual camera, $\mathbf{P}_v$;
5. generate the virtual view as $\mathbf{x}_v^{(k)} \cong \mathbf{P}_v\mathbf{X}_k^P$.

In spite of its simplicity, this algorithm has a number of tricky points. First of all, as noted by Fusiello in [8, 4], a crucial step is point 5, namely the need to specify the position of the virtual camera in the projective frame. In general, we are interested in synthesizing views by moving the virtual camera along a convenient smooth path between the two cameras. While in the Euclidean space it is easy to specify a smooth trajectory, this is not necessarily true in the projective frame. Hence, a proper method to specify the virtual camera in the projective frame is required. Moreover, as can be noticed by looking at the work by Tebaldini [22], the knowledge of the homography $\mathbf{A}_{ij}(\pi)$ induced by a general reference plane $\boldsymbol{\pi}$ is not sufficient for solving the view synthesis problem. In fact, a number of Euclidean quantities remain undetermined [22] and have to be estimated by exploiting further geometric constraints arising from the two views.

The solution to the first problem relies on the concepts of Lie Algebra and Lie Groups described in the next section. Regarding the second issue, in Section 5 we will show how the knowledge of the infinite homography is a sufficient condition for the synthesis of new views. Formally, this corresponds to choosing the plane at infinity as a reference. If the plane at infinity is identifiable, one may proceed as in [8]; however, in most of the cases the images don't show the plane at infinity, preventing the use of this solution. Several strategies for estimating the infinite homography, not based on the search of the plane at the infinity, have been recently proposed in the literature, as for example in [24],[13]. Furthermore, at the end of Section 5 we also propose two novel methods for determining the infinite homography exploiting the relative affine structure jointly with geometric elements visible in the two views.

4. Linear Lie Groups

In this section we introduce some concepts related to the Lie Groups and the Lie Algebra. They represent powerful tools for the computation of smooth trajectories both in the Euclidean space and in the Projective space. The basic idea is that a rigid motion can be synthesized through a linear combination of infinitesimal generators. This results in a smooth trajectory that interpolates the initial and final positions of an object.

4.1. Special Euclidean Group and rigid motions

Consider a transformation matrix given by

$$\bar{\mathbf{G}} = \begin{bmatrix} \mathbf{R} & \mathbf{t} \\ \mathbf{0}^T & 1 \end{bmatrix} , \tag{13}$$

where $\mathbf{R}$ is a 3D rotation matrix and $\mathbf{t} \in \mathbb{R}^3$ is a translation vector. The set of all the transformations in the form of Equation (13) defines the set of all the rigid motions in $\mathbb{R}^3$, characterized by 6 degrees of freedom (3 for the rotation and 3 for the translation). Under the matrix multiplication operation, this set identifies the so-called *Special Euclidean Group* of $\mathbb{R}^3$, indicated by $SE(3)$. A group is a pair formed by a set and a binary operation which satisfies the properties of closeness, associativity, existence and uniqueness of the identity and of the inverse elements. If a group is also a differentiable manifold it is a *Lie group*, and it represents a generalization of the concepts of differentiable curves. It can be shown that $SE(3)$ is a Lie Group [9]. Consider now the parametrization of the matrix in Equation (13) given by $\mathbf{G}(\boldsymbol{\theta}) \in SE(3)$, where $\boldsymbol{\theta} = [\theta_1, \ldots, \theta_6]^T$ is the parameter vector. A curve in the parameter space is defined as $\gamma : \theta_i = \theta_i(t)$, where t is a real parameter. Alternatively, the curve can be defined on the manifold, namely $\Gamma : \mathbf{G} = \mathbf{G}(\boldsymbol{\theta}(t))$. Without loss of generality we can assume that the curve Γ on the manifold starts from the identity, i.e. $\mathbf{G}(\boldsymbol{\theta}(0)) = \mathbf{I}$. Since $SE(3)$ is differentiable, we may compute the derivative of $\mathbf{G}$ along the curve

$$\frac{d\mathbf{G}}{dt} = \sum_{i=1}^{6} \frac{d\theta_i}{dt} \frac{\partial \mathbf{G}}{\partial \theta_i} .$$

We now introduce the concept of *Lie algebra*, which is the set of the differentials in $t = 0$ of all the differentiable curves Γ on the manifold such that $\Gamma(0) = \mathbf{I}$. The Lie algebra of $SE(3)$, denoted by $se(3)$, may therefore be interpreted as the tangent space of the manifold at the identity.

More specifically, $\mathbf{G}$ represents a point on the manifold and $d\mathbf{G}/dt$ is a vector tangent to the curve Γ at the identity. Moving along Γ by small step dt, we may write

$$\mathbf{G}(dt) = \mathbf{I} + dt \frac{d\mathbf{G}}{dt} ,$$

that may be interpreted as a displacement on the manifold due to the right multiplication by the element $\mathbf{I} + dt\frac{d\mathbf{G}}{dt}$, namely:

$$\mathbf{G}(dt) = \mathbf{I} \left(\mathbf{I} + dt \frac{d\mathbf{G}}{dt} \right) .$$

At this point, we could think to replicate this displacement K times, yielding

$$\mathbf{G}(Kdt) = \left(\mathbf{I} + dt \frac{d\mathbf{G}}{dt} \right)^K .$$

As a consequence, by posing $t = Kdt$ and letting K go to infinity, we obtain

$$\mathbf{G}(t) = \lim_{K \to \infty} \left(\mathbf{I} + \frac{t}{K} \frac{d\mathbf{G}}{dt} \right)^K = e^{t \frac{d\mathbf{G}}{dt}} . \tag{14}$$

By posing $t = 1$ Equation (14) defines a map between the elements of $SE(3)$ and those of the correspondent Lie algebra $se(3)$ in the form $\mathbf{G} = e^{\frac{d\mathbf{G}}{dt}}$. This mapping is called the *Exponential Map*. Returning to the geometrical meaning of rigid motions of the elements of $SE(3)$, we can see that, as t varies over the real axis, Equation (14) describes a one parameter family of transformations, emanating from the identity and passing through $\bar{\mathbf{G}}$ at $t = 1$. In other words, Equation (14) defines a method for interpolating and extrapolating new transformations on the basis of an element of the Lie algebra, $\frac{d\mathbf{G}}{dt}$, that may be thought as the infinitesimal generator of the transformation defined by $\mathbf{G}$.

The question now is how to obtain the infinitesimal generator from $\mathbf{G}$. In general, the correspondence between a Lie group and its Lie algebra is not one to one, preventing the uniqueness of the interpolated (extrapolated) motions. This issue is solved by a theorem by E. Cartan [10], assuring that a one to one correspondence exists between a Lie group and its Lie algebra provided that the Lie group is simply connected. Dealing with matrix groups, this condition may be reduced to require that no eigenvalue of $\mathbf{G}$ lies on the closed negative real axis [8]. Under this condition the infinitesimal generator of $\mathbf{G}$ may be computed as the principal matrix logarithm of $\bar{\mathbf{G}}$:

$$\frac{d\mathbf{G}}{dt} = \log(\bar{\mathbf{G}}) = -\sum_{K=1}^{\infty} \frac{(\mathbf{I} - \bar{\mathbf{G}})^K}{K} .$$

Therefore, the synthesized motion may be obtained as

$$\mathbf{G}(t) = e^{t \log(\bar{\mathbf{G}})} = \bar{\mathbf{G}}^t . \tag{15}$$

Since this condition is always satisfied for the elements of $SE(3)$ [8], what is left to show prior to use Equation (14) in the applications is that the interpolated (extrapolated) transformations are actually smooth, in some sense. To do this, it is sufficient to note that:

- if the motion represented by $\bar{\mathbf{G}}$ is a rotation by an angle ϕ around an axis, then $\bar{\mathbf{G}}^t$ represents a rotation by an angle $t\phi$ around the same axis (as stated by the Rodrigues' formula);
- if the motion represented by $\bar{\mathbf{G}}$ is a translation by a vector $\mathbf{t}$, then $\bar{\mathbf{G}}^t$ represents a translation by a vector $\mathbf{t}$.

Hence, the motions generated by Equation (14) are smooth in the sense specified by [1] and [17].

4.2. Lie group homomorphism

So far, it has been shown that the exponential map provides a tool for generating rigid motions in $SE(3)$. Dealing with rigid motions, however, does not suffice for the aim of view synthesis from uncalibrated reference views. In fact, we will see that the Euclidean frame is inaccessible, and thus the matrix $\bar{\mathbf{G}}$ is not known. The key to solve

this problem is represented by the concept of Lie group homomorphism [9].

Given two Lie groups Ω_G and Ω_T, a Lie group homomorphism is an application $\varphi : \Omega_G \to \Omega_T$ such that

$$\varphi(\mathbf{G}_1\mathbf{G}_2) = \varphi(\mathbf{G}_1)\varphi(\mathbf{G}_2) = \mathbf{T}_1\mathbf{T}_2$$

for all $\mathbf{G}_1, \mathbf{G}_2 \in \Omega_G$, where $\mathbf{T}_1, \mathbf{T}_2 \in \Omega_T$. From this definition it follows that $\varphi(\mathbf{I}) = \mathbf{I}$, in fact

$$\varphi(\mathbf{G}) = \varphi(\mathbf{I})\varphi(\mathbf{G}) \ ,$$

and from the uniqueness of the identity one gets that $\varphi(\mathbf{I}) = \mathbf{I}$.

Consider now the application of φ to Equation (14). Under the assumption that $\varphi(\mathbf{G})$ is also differentiable, we may write

$$\begin{aligned} \mathbf{T}(t) &= \varphi(\mathbf{G}(t)) \\ &= \varphi\left(\lim_{K\to\infty}\left(\mathbf{I} + \frac{t}{K}\frac{d\mathbf{G}}{dt}\right)^K\right) \\ &= \lim_{K\to\infty}\varphi\left(\left(\mathbf{I} + \frac{t}{K}\frac{d\mathbf{G}}{dt}\right)^K\right) \\ &= \lim_{K\to\infty}\left(\varphi\left(\mathbf{I} + \frac{t}{K}\frac{d\mathbf{G}}{dt}\right)\right)^K \ . \end{aligned} \tag{16}$$

Neglecting higher order terms, the action of φ may be approximated as

$$\begin{aligned} \varphi\left(\mathbf{I} + \frac{t}{K}\frac{d\mathbf{G}}{dt}\right) &= \mathbf{I} + \frac{t}{K}\sum_k\left[\frac{\partial\varphi}{\partial g_k}\right]_{\mathbf{I}}\frac{dg_k}{dt} \\ &= \mathbf{I} + \frac{t}{K}d\varphi\left(\frac{d\mathbf{G}}{dt}\right) \ , \end{aligned} \tag{17}$$

where $d\varphi$ defines an application between the tangent spaces at the identity of Ω_G and Ω_T induced by φ. This operation, also called derivative of φ, maps the derivative of $\mathbf{G}(t)$ on the derivative of $\mathbf{T}(t) = \varphi(\mathbf{G}(t))$. In fact, a direct evaluation of the derivative of $\mathbf{T}(t)$ results in

$$\begin{aligned} \frac{d\mathbf{T}}{dt} &= \frac{d\varphi}{dt} \\ &= \sum_k\left[\frac{\partial\varphi}{\partial g_k}\right]_{\mathbf{I}}\frac{dg_k}{dt} = d\varphi\left(\frac{d\mathbf{G}}{dt}\right) \end{aligned} \tag{18}$$

that may be plugged into Equation (16) along with Equation (17), yielding

$$\mathbf{T}(t) = \lim_{K\to\infty}\left(\mathbf{I} + \frac{t}{K}\frac{d\mathbf{T}}{dt}\right)^K = e^{t\frac{d\mathbf{T}}{dt}} \ . \tag{19}$$

Hence, we obtain that under the action of a Lie group homomorphism, the following diagram commute:

$$\begin{array}{ccc} \mathbf{G}^t & \xrightarrow{\varphi} & \mathbf{T}^t \\ \uparrow e^{t\cdot()} & & \uparrow e^{t\cdot()} \\ \dfrac{d\mathbf{G}}{dt} & \xrightarrow{d\varphi} & \dfrac{d\mathbf{T}}{dt} \ . \end{array} \tag{20}$$

4.3. The Special Projective Group

As we will see in the next section, the whole problem of view synthesis through linear combinations of elements of the Lie algebra may be essentially reduced to finding a homomorphism for the transformations between the cameras in the projective and in the Euclidean frames. For this reason we introduce here the concept of Special Projective Group.

Consider a 3×3 invertible upper triangular matrix $\mathbf{K}$ and a matrix $\mathbf{G} \in SE(3)$. By posing

$$\tilde{\mathbf{K}} = \begin{bmatrix} \mathbf{K} & \mathbf{0} \\ \mathbf{0} & 1 \end{bmatrix} \ ,$$

it is easy to verify that the transformation matrix

$$\mathbf{T} = \tilde{\mathbf{K}}\mathbf{G}\tilde{\mathbf{K}}^{-1} \tag{21}$$

is homomorphic to $\mathbf{G}$. In fact, given 2 matrices $\mathbf{G}_1, \mathbf{G}_2 \in SE(3)$, we have that

$$\begin{aligned} \mathbf{T}_{12} &= \tilde{\mathbf{K}}\mathbf{G}_1\tilde{\mathbf{K}}^{-1}\tilde{\mathbf{K}}\mathbf{G}_2\tilde{\mathbf{K}}^{-1} \\ &= \tilde{\mathbf{K}}\mathbf{G}_1\mathbf{G}_2\tilde{\mathbf{K}}^{-1} \\ &= \tilde{\mathbf{K}}\mathbf{G}_{12}\tilde{\mathbf{K}}^{-1} \ , \end{aligned}$$

where $\mathbf{G}_{12} = \mathbf{G}_1\mathbf{G}_2 \in SE(3)$. Hence, by virtue of (20), $\mathbf{T}^t$ is homomorphic to $\mathbf{G}^t$ for every t, i.e.

$$\mathbf{T}^t = \tilde{\mathbf{K}}\mathbf{G}^t\tilde{\mathbf{K}}^{-1} \ .$$

The matrices in the form of Equation (21) are the members of the so-called *Special Projective Group*, denoted by $SP(3)$. In order to account the scalability of these matrices, $SP(3)$ is defined as the quotient set of the matrices in the form of Equation (21) with respect to the equivalence relation $\cong$.

In the next section we will show that the transformation between two camera matrices with identical calibration matrix $\mathbf{K}$ belongs to the group $SP(3)$, having the form of the matrix $\mathbf{T}$. The synthesized transformation $\mathbf{T}^t$ is therefore perfectly equivalent to the transformation yielded by removing the camera matrix, applying a proper rotation and translation, and finally reapplying the camera matrix, even in absence of information about the calibration parameters.

5. View synthesis

An effective way to describe the relative displacement and orientation of two cameras is that of considering the transformation the scene should undergo for the same views to be generated by a single camera taking a picture of the scene before and after such transformation. For this purpose, we observe that, given the camera matrices $\mathbf{P}_i = \mathbf{K}_i[\mathbf{R}_i|\mathbf{t}_i]$ and $\mathbf{P}_j = \mathbf{K}_j[\mathbf{R}_j|\mathbf{t}_j]$, the projections of the point $\mathbf{X}$ can be rewritten as

$$\begin{aligned} \mathbf{x}_i &\cong [\mathbf{I}|\mathbf{0}]\begin{bmatrix} \mathbf{P}_i & \\ \mathbf{0}^T & 1 \end{bmatrix}\mathbf{X} = [\mathbf{I}|\mathbf{0}]\,\mathbf{T}_i\mathbf{X} = [\mathbf{I}|\mathbf{0}]\,\mathbf{X}_i \\ \mathbf{x}_j &\cong [\mathbf{I}|\mathbf{0}]\begin{bmatrix} \mathbf{P}_j & \\ \mathbf{0}^T & 1 \end{bmatrix}\mathbf{X} = [\mathbf{I}|\mathbf{0}]\,\mathbf{T}_j\mathbf{X} = [\mathbf{I}|\mathbf{0}]\,\mathbf{X}_j \ , \end{aligned}$$

where $\mathbf{T}_i$ ($\mathbf{T}_j$) is the transformation of $\mathbb{P}^3$ mapping the scene to the frame centred on the i-th (j-th) camera. Since the cameras are finite, the matrices $\mathbf{T}_i$ and $\mathbf{T}_j$ are invertible; by posing $\mathbf{T}_{ij} = \mathbf{T}_j\mathbf{T}_i^{-1}$ we obtain that $\mathbf{X}_j = \mathbf{T}_{ij}\mathbf{X}_i$ and

$$\mathbf{x}_j \cong [\mathbf{I}|\mathbf{0}]\,\mathbf{T}_{ij}\mathbf{X}_i \ .$$

The inverse of $\mathbf{T}_i$ can be explicitly calculated, resulting in

$$\mathbf{T}_i^{-1} = \begin{bmatrix} \mathbf{R}_i^{-1}\mathbf{K}_i^{-1} & -\mathbf{R}_i^{-1}\mathbf{t}_i \\ \mathbf{0}^T & 1 \end{bmatrix} \ ;$$

therefore we also have that

$$\mathbf{T}_{ij} = \begin{bmatrix} \mathbf{K}_j\mathbf{R}_j\mathbf{R}_i^{-1}\mathbf{K}_i^{-1} & -\mathbf{K}_j\mathbf{R}_j\mathbf{R}_i^{-1}\mathbf{t}_i + \mathbf{K}_j\mathbf{t}_j \\ \mathbf{0}^T & 1 \end{bmatrix} \ .$$

The top left part of $\mathbf{T}_{ij}$ is exactly the definition of the infinite homography given in Equation (2); furthermore, considering that the centre of the first camera is $-\mathbf{R}_i^{-1}\mathbf{t}_i$, its projection on the second view is the epipole

$$\mathbf{e}_{ji} \cong \bar{\mathbf{e}}_{ji} = \mathbf{K}_j[\mathbf{R}_j|\mathbf{t}_j] \begin{bmatrix} -\mathbf{R}_i^{-1}\mathbf{t}_i \\ 1 \end{bmatrix} \ .$$

As a consequence, the transformation mapping the scene from view i to view j can be rewritten in the following interesting form:

$$\mathbf{T}_{ij} = \begin{bmatrix} \mathbf{A}_{ij}^{\infty} & \bar{\mathbf{e}}_{ji} \\ \mathbf{0}^T & 1 \end{bmatrix} \ .$$

If the two cameras have identical calibration matrices, i.e. $\mathbf{K} = \mathbf{K}_i = \mathbf{K}_j$, then $\mathbf{T}_{ij}$ belongs to the Special Projective Group $SP(3)$. Therefore, in principle, a virtual view can be generated by exponentiating the matrix $\mathbf{T}_{ij}$:

$$\mathbf{x}(t) \cong [\mathbf{I}|\mathbf{0}]\mathbf{T}_{ij}^t\mathbf{X}_i \ . \tag{22}$$

The problem is that the actual reconstruction $\mathbf{X}_i$ is unknown, since only a projective reconstruction of the scene is available.

Following the algorithm described in Section 3, the projective reconstruction is given by

$$\mathbf{X}_i^P = [\mathbf{x}_i^T, \mu_i]^T \ ,$$

where μ_i is the relative affine structure with respect to the first camera. When the plane at infinity is chosen as a reference, we saw that the relative affine structure is given by

$$\mu_i = \frac{z_i^0}{z_i} \ .$$

Owing to the projective reconstruction ambiguity, for some homography $\mathbf{H}$ we also have that

$$\begin{aligned} \mathbf{X}_i^P &\cong \mathbf{H}\mathbf{X} \\ &= \mathbf{H}\mathbf{T}_i^{-1}\mathbf{X}_i \\ &= \mathbf{H}_i\mathbf{X}_i \ . \end{aligned} \tag{23}$$

By posing $\mathbf{x}_i = [x_i, y_i, 1]^T$, the point $\mathbf{X}_i$ has Euclidean coordinates $z_i\mathbf{x}_i$, i.e. $\mathbf{X}_i = [z_ix_i, z_iy_i, z_i, 1]^T$ and, equivalently

$$z_i\mathbf{x}_i = [\mathbf{I}|\mathbf{0}]\mathbf{X}_i \ .$$

At this point, from Equation (23) it is immediate to see that

$$\mathbf{H}_i \cong \begin{bmatrix} \mathbf{I} & \mathbf{0} \\ \mathbf{0}^T & z_i^0 \end{bmatrix} \ .$$

The same reasoning leads to the expression of the projectivity $\mathbf{H}_j$ between the scene seen from the point of view of the second camera $\mathbf{X}_j = \mathbf{T}_{ij}\mathbf{X}_i$, and its reconstruction $\mathbf{X}_j^P$. This is given by

$$\mathbf{H}_j \cong \begin{bmatrix} \mathbf{I} & \mathbf{0} \\ \mathbf{0}^T & z_j^0 \end{bmatrix} \ .$$

Notice now that the projections of $\mathbf{X}_i^P$ on the two views are obtained as

$$\begin{aligned} \mathbf{x}_i &\cong [\mathbf{I}|\mathbf{0}]\mathbf{X}_i^P \\ \mathbf{x}_j &\cong [\mathbf{I}|\mathbf{0}]\mathbf{X}_j^P = [\mathbf{I}|\mathbf{0}]\mathbf{D}_{ij}\mathbf{X}_i^P \ , \end{aligned}$$

where $\mathbf{D}_{ij}$ is the relative transformation between the cameras as seen in the projective frame. In order to obtain its expression it is sufficient to write

$$\mathbf{X}_j^P \cong \mathbf{H}_j\mathbf{X}_j = \mathbf{H}_j\mathbf{T}_{ij}\mathbf{X}_i \cong \mathbf{H}_j\mathbf{T}_{ij}\mathbf{H}_i^{-1}\mathbf{X}_i^P \ ,$$

and therefore

$$\mathbf{D}_{ij} = \mathbf{H}_j\mathbf{T}_{ij}\mathbf{H}_i^{-1} = \begin{bmatrix} \mathbf{A}_{ij}^{\infty} & \frac{1}{z_i^0}\bar{\mathbf{e}}_{ji} \\ \mathbf{0}^T & \frac{z_j^0}{z_i^0} \end{bmatrix} \ . \tag{24}$$

Supposing that the infinite homography $\mathbf{A}_{ij}^{\infty}$ is known, this matrix is fully determined by observing the following reasoning. Consider the representations of the reference point $\mathbf{X}^0$ in the reference systems of the two cameras, i.e.

$$\begin{aligned} \mathbf{X}_i^0 &= \mathbf{T}_i\mathbf{X}^0 = [x_i^0, y_i^0, z_i^0, 1]^T \\ \mathbf{X}_j^0 &= \mathbf{T}_j\mathbf{X}^0 = [x_j^0, y_j^0, z_j^0, 1]^T \ . \end{aligned}$$

Since $\mathbf{X}_j^0 = \mathbf{T}_{ij}\mathbf{X}_i^0$, it is easy to verify that

$$\frac{z_j^0}{z_i^0}\mathbf{x}_j^0 = \mathbf{A}_{ij}^{\infty}\mathbf{x}_i^0 + \frac{1}{z_i^0}\bar{\mathbf{e}}_{ji} \ , \tag{25}$$

where $\mathbf{x}_i^0 = [x_i^0/z_i^0, y_i^0/z_i^0, 1]^T$ and $\mathbf{x}_j^0 = [x_j^0/z_j^0, y_j^0/z_j^0, 1]^T$ are the projections of the reference point on the two views. On the other hand, for the relative affine structure property we have that

$$\begin{aligned} \mathbf{x}_j^0 &\cong \mathbf{A}_{ij}^{\infty}\mathbf{x}_i^0 + \mathbf{e}_{ji} \\ \Longrightarrow \beta\mathbf{x}_j^0 &= \mathbf{A}_{ij}^{\infty}\mathbf{x}_i^0 + \mathbf{e}_{ji} \ , \end{aligned} \tag{26}$$

where $\mathbf{e}_{ji}$ is scaled in such a way to have $\mu_i = 1$ for $\mathbf{x}_i^0$, and $\beta \neq 0$ is an opportune scaling factor. Since all the terms in Equation (26) are known, β is readily determined. Comparing eqs. (25) and (26) we notice that they are

scaled in the same way, and therefore we finally obtain that

$$\frac{z_j^0}{z_i^0} = \beta, \quad \frac{1}{z_i^0}\bar{\mathbf{e}}_{ji} = \mathbf{e}_{ji} \ . \tag{27}$$

An alternate way for estimating the matrix $\mathbf{D}_{ij}$ is to use a sufficient number of correspondent pairs $(\mathbf{X}_i^P, \mathbf{X}_j^P)$. In this case, the scaling ambiguity of the estimated matrix can be solved by exploiting the fact that $\det(\mathbf{A}_{ij}^\infty) = 1$.

5.1. The scene calibration matrix

In principle, the knowledge of the transformation matrix $\mathbf{D}_{ij}$ could be exploited to synthesize the virtual view as

$$\mathbf{x}(t) \cong [\mathbf{I}|\mathbf{0}]\mathbf{D}_{ij}^t\mathbf{X}_i^P$$

as explained in Section 4. What prevents from the application of this method is that, in general, $\mathbf{D}_{ij}$ is not homomorphic to $\mathbf{T}_{ij}$, resulting in a loss of physical validity of the synthesized view. To overcome this issue, we observe that a sufficient and necessary condition for a homomorphism to be between $\mathbf{D}_{ij}$ and $\mathbf{T}_{ij}$ is that the two matrices are similar[3]. This condition may be imposed by introducing a matrix $\mathbf{V}_j$ such that $\mathbf{V}_j\mathbf{D}_{ij}$ is similar to $\mathbf{T}_{ij}$. From Equation (24) it follows that such a matrix is obtained as

$$\mathbf{V}_j = \mathbf{H}_i\mathbf{H}_j^{-1} = \begin{bmatrix} \mathbf{I} & \mathbf{0} \\ \mathbf{0}^T & \frac{z_i^0}{z_j^0} \end{bmatrix} = \begin{bmatrix} \mathbf{I} & \mathbf{0} \\ \mathbf{0}^T & \beta \end{bmatrix} \ .$$

Hence

$$\mathbf{V}_j\mathbf{D}_{ij} = \begin{bmatrix} \mathbf{A}_{ij}^\infty & \mathbf{e}_{ji} \\ \mathbf{0}^T & 1 \end{bmatrix} \ . \tag{28}$$

where $\mathbf{e}_{ji}$ is scaled in such a way to have $\mu_i = 1$ for $\mathbf{x}_i^0$.

Since $\mathbf{V}_j\mathbf{D}_{ij}$ is homomorphic to $\mathbf{T}_{ij}$, it is easily verified that

$$(\mathbf{V}_j\mathbf{D}_{ij})^t = \mathbf{H}_i\mathbf{T}_{ij}^t\mathbf{H}_i^{-1} \quad \forall\, t \in \mathbb{R} \ ,$$

and consequently the novel view is synthesized exactly as if the transformation $\mathbf{T}_{ij}$ (and hence the relative motion $\mathbf{G}_{ij}$) were known:

$$\begin{aligned} \mathbf{x}(t) &\cong [\mathbf{I}|\mathbf{0}](\mathbf{V}_j\mathbf{D}_{ij})^t\mathbf{X}_i^P \\ &= [\mathbf{I}|\mathbf{0}]\mathbf{H}_i\mathbf{T}_{ij}^t\mathbf{X}_i \\ &= [\mathbf{I}|\mathbf{0}]\mathbf{T}_{ij}^t\mathbf{X}_i \ . \end{aligned} \tag{29}$$

The matrix $\mathbf{V}_j$ acts therefore as an equalizer between the two reconstructions, balancing the projective deformations in such a way as to reach similarity to $\mathbf{T}_{ij}$. For this reason, $\mathbf{V}_j$ is called *scene calibration matrix*.

[3]Two square matrices $\mathbf{A}$ and $\mathbf{B}$ are similar if $\mathbf{B} = \mathbf{P}^{-1}\mathbf{A}\mathbf{P}$ for some non-singular matrix $\mathbf{P}$.

5.2. Calculating the infinite homography

In the last section we saw that the knowledge of the infinite homography $\mathbf{A}_{ij}^\infty$ is sufficient for calculating the matrices $\mathbf{D}_{ij}$ and $\mathbf{V}_j$, and it is therefore a sufficient condition for synthesizing virtual views from a pair of uncalibrated reference views. The focus is now on the methodologies for estimating the transformation induced by the plane at infinity.

The straightforward approach for estimating the infinite homography is based on the knowledge of 3 vanishing points, i.e. the projections of 3 points at infinity. The infinite homography is therefore estimated as the solution of the system

$$\begin{cases} \mathbf{p}_j^{(1)} & \cong \mathbf{A}_{ij}^\infty\mathbf{p}_i^{(1)} \\ \mathbf{p}_j^{(2)} & \cong \mathbf{A}_{ij}^\infty\mathbf{p}_i^{(2)} \\ \mathbf{p}_j^{(3)} & \cong \mathbf{A}_{ij}^\infty\mathbf{p}_i^{(3)} \\ \mathbf{e}_{ji} & \cong \mathbf{A}_{ij}^\infty\mathbf{e}_{ij} \end{cases} \ ,$$

where $(\mathbf{p}_i^{(k)}, \mathbf{p}_j^{(k)})$ with $k = 1, \ldots, 3$ refers to the 3 pairs of correspondent vanishing points. In general, this method is rarely applicable because the images don't show elements that allow to calculate more than one or two vanishing points. Moreover, the estimation of a vanishing point in some cases can be very difficult, especially when it is retrieved by intersecting two lines that are almost parallel, leading to high numerical errors.

In the literature many methods for estimating the infinite homography were recently proposed. In [24] the authors show how the infinite homography can be retrieved starting from two arbitrary rectangles visible in the two views, exploiting the so-called semi-metric transformation. An alternative method is proposed in [13], where the infinite homography is linearly estimated from the epipolar constraints.

Here we propose two novel methods for estimating the infinite homography which exploit its relation with the homography induced by a reference plane $\boldsymbol{\pi}$ given in Equation (9). This relation reveals a scaling ambiguity between the left and right side. However, exploiting some of the results obtained in the previous sections, it is possible to solve the ambiguity; furthermore, we will see that such a scaling is independent from the choice of the reference plane. Equation (9) can be rewritten as follows:

$$\mathbf{A}_{ij}(\pi) = k_\pi\left(\mathbf{A}_{ij}^\infty - \bar{\mathbf{e}}_{ji}\frac{\mathbf{v}_i^T}{c_i}\right) \ ,$$

or equivalently as

$$\mathbf{A}_{ij}^\infty = \frac{1}{k_\pi}\mathbf{A}_{ij}(\pi) + \bar{\mathbf{e}}_{ji}\frac{\mathbf{v}_i^T}{c_i} \ , \tag{30}$$

where $k_\pi \neq 0$ is the unknown scaling factor. Another important result is given by Equation (25), which shows that the relation between the points in the two views by means

of the infinite homography $\mathbf{A}_{ij}^{\infty}$ and its relative affine structure is not affected by scaling ambiguity. Plugging Equation (30) into Equation (25) we obtain

$$\begin{aligned}\frac{z_j^0}{z_i^0}\mathbf{x}_j^0 &= \left(\frac{1}{k_\pi}\mathbf{A}_{ij}(\pi)+\bar{\mathbf{e}}_{ji}\frac{\mathbf{v}_i^T}{c_i}\right)\mathbf{x}_i+\frac{1}{z_i^0}\bar{\mathbf{e}}_{ji}\\ &= \frac{1}{k_\pi}\mathbf{A}_{ij}(\pi)\mathbf{x}_i+\bar{\mathbf{e}}_{ji}\left(\frac{\mathbf{v}_i^T}{c_i}\mathbf{x}_i+\frac{1}{z_i^0}\right)\\ &= \frac{1}{k_\pi}\mathbf{A}_{ij}(\pi)\mathbf{x}_i+\bar{\mathbf{e}}_{ji}\left(\frac{\boldsymbol{\pi}_i^T\mathbf{X}_i^0}{z_i^0 c_i}\right)\ , \end{aligned} \quad (31)$$

where $\boldsymbol{\pi}_i=[\mathbf{v}_i^T,c_i]^T$. Notice that, in general, the epipole is known up to a scale factor, i.e. $\gamma\bar{\mathbf{e}}_{ji}=\mathbf{e}_{ji}$ for some $\gamma\neq 0$. Taking into account this fact, multiplying both terms of Equation (31) by $\gamma\frac{\boldsymbol{\pi}_i^T\mathbf{X}_i^0}{z_i^0 c_i}$ we obtain

$$\gamma\left(\frac{z_j^0}{z_i^0}\frac{z_i^0 c_i}{\boldsymbol{\pi}_i^T\mathbf{X}_i^0}\right)\mathbf{x}_j^0=\gamma\frac{z_i^0 c_i}{\boldsymbol{\pi}_i^T\mathbf{X}_i^0}\frac{1}{k_\pi}\mathbf{A}_{ij}(\pi)\mathbf{x}_i^0+\mathbf{e}_{ji}\ , \quad (32)$$

from which we recognize the property of the relative affine structure, scaled in such a way that $\mu_i=1$ for the reference point $\mathbf{x}_i^0$, i.e.

$$b_\pi\mathbf{x}_j^0=\mathbf{A}_{ij}(\pi)\mathbf{x}_i^0+\mathbf{e}_{ji} \quad (33)$$

for some $b_\pi\neq 0$. Comparing Equation (32) with Equation (33) it is immediate to see that

$$\begin{cases}\gamma\frac{z_i^0 c_i}{\boldsymbol{\pi}_i^T\mathbf{X}_i^0}\frac{1}{k_\pi}=1\\ b_\pi=\gamma\left(\frac{z_j^0}{z_i^0}\frac{z_i^0 c_i}{\boldsymbol{\pi}_i^T\mathbf{X}_i^0}\right)\end{cases}\Rightarrow k_\pi=b_\pi\frac{z_i^0}{z_j^0}\ .$$

At this point, we can rewrite Equation (30) as follows

$$\mathbf{A}_{ij}^{\infty}=\frac{1}{b_\pi}\frac{z_j^0}{z_i^0}\mathbf{A}_{ij}(\pi)+\bar{\mathbf{e}}_{ji}\frac{\mathbf{v}_i^T}{c_i}\ ,$$

or equivalently

$$\frac{1}{b_\pi}\mathbf{A}_{ij}(\pi)=\frac{z_i^0}{z_j^0}\left(\mathbf{A}_{ij}^{\infty}-\bar{\mathbf{e}}_{ji}\frac{\mathbf{v}_i^T}{c_i}\right)\ .$$

Notice that, if $\mathbf{A}_{ij}(\pi)$ and $\mathbf{e}_{ji}$ are known, they can always be scaled in such a way that Equation (33) becomes

$$\mathbf{x}_j^0=\mathbf{A}_{ij}^{\star}(\pi)\mathbf{x}_i^0+\mathbf{e}_{ji}^{\star}\ ,$$

where $\mathbf{A}_{ij}^{\star}(\pi)=\mathbf{A}_{ij}(\pi)/b_\pi$ and $\mathbf{e}_{ji}^{\star}=\mathbf{e}_{ji}/b_\pi$. According to this scaling, we final obtain a fundamental relation between the infinite homography and the homography induced by a reference plane $\boldsymbol{\pi}$, which is independent from the choice of the plane:

$$\mathbf{A}_{ij}^{\star}(\pi)=\frac{z_i^0}{z_j^0}\left(\mathbf{A}_{ij}^{\infty}-\bar{\mathbf{e}}_{ji}\frac{\mathbf{v}_i^T}{c_i}\right)\ . \quad (34)$$

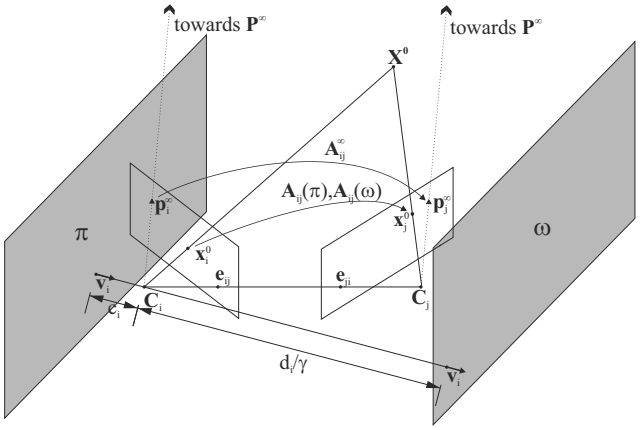


Figure 3: Estimation of the Infinite Homography $\mathbf{A}_{ij}^{\infty}$ from a pair of parallel planes and a vanishing point not on the planes$\mathbb{P}^{\infty}$.

5.2.1. Infinite homography from a pair of parallel planes and a vanishing point not on the planes

Suppose that in the two views it is possible to identify a pair of parallel planes, whose coordinates are $\boldsymbol{\pi}_i=[\mathbf{v}_i^T,c_i]^T$ and $\boldsymbol{\omega}_i=[\gamma\mathbf{v}_i^T,d_i]^T$ in the reference system of the first camera, with $\gamma\in\mathbb{R}-\{0,1\}$ (see Figure 3). According to Equation (34), for the two planes we have

$$\begin{aligned}\mathbf{A}_{ij}^{\star}(\pi) &= \frac{z_i^0}{z_j^0}\left(\mathbf{A}_{ij}^{\infty}-\bar{\mathbf{e}}_{ji}\frac{\mathbf{v}_i^T}{c_i}\right) \quad (35)\\ \mathbf{A}_{ij}^{\star}(\omega) &= \frac{z_i^0}{z_j^0}\left(\mathbf{A}_{ij}^{\infty}-\bar{\mathbf{e}}_{ji}\frac{\gamma\mathbf{v}_i^T}{d_i}\right)\ .\end{aligned}$$

Subtracting the two equations we get

$$\mathbf{A}_{ij}^{\star}(\pi)-\mathbf{A}_{ij}^{\star}(\omega)=\frac{z_i^0}{z_j^0}\bar{\mathbf{e}}_{ji}\alpha\frac{\mathbf{v}_i^T}{d_i}\ ,$$

where $\alpha=\frac{\gamma c_i}{d_i}-1$. By left-multiplying both terms by $\frac{\bar{\mathbf{e}}_{ji}^T}{||\bar{\mathbf{e}}_{ji}||^2}$ we obtain

$$\frac{\mathbf{v}_i^T}{c_i}=\frac{z_j^0}{z_i^0}\frac{1}{\alpha}\frac{\bar{\mathbf{e}}_{ji}^T}{||\bar{\mathbf{e}}_{ji}||^2}\left(\mathbf{A}_{ij}^{\star}(\pi)-\mathbf{A}_{ij}^{\star}(\omega)\right)\ . \quad (36)$$

Inserting Equation (36) into Equation (35) and rearranging the terms leads to the following expression for the infinite homography

$$\mathbf{A}_{ij}^{\infty}=\frac{z_j^0}{z_i^0}\left[\mathbf{A}_{ij}^{\star}(\pi)+\frac{1}{\alpha}\frac{\bar{\mathbf{e}}_{ji}\bar{\mathbf{e}}_{ji}^T}{||\bar{\mathbf{e}}_{ji}||^2}\left(\mathbf{A}_{ij}^{\star}(\pi)-\mathbf{A}_{ij}^{\star}(\omega)\right)\right], \quad (37)$$

where the ratio $\frac{z_i^0}{z_j^0}$ and the constant α are unknown. Notice that the matrix $\frac{\bar{\mathbf{e}}_{ji}\bar{\mathbf{e}}_{ji}^T}{||\bar{\mathbf{e}}_{ji}||^2}$ is independent from the scale of the epipole, i.e.

$$\frac{\bar{\mathbf{e}}_{ji}\bar{\mathbf{e}}_{ji}^T}{||\bar{\mathbf{e}}_{ji}||^2}=\frac{\lambda\bar{\mathbf{e}}_{ji}\lambda\bar{\mathbf{e}}_{ji}^T}{||\lambda\bar{\mathbf{e}}_{ji}||^2}$$

for any $\lambda\neq 0$, and therefore this term is unequivocally determined from the knowledge of $\mathbf{e}_{ji}$. The unknowns can

be determined by considering a pair of correspondent vanishing points $(\mathbf{p}_i^\infty, \mathbf{p}_j^\infty)$ in the two views, not lying on the planes. In fact, by definition

$$\mathbf{p}_j^\infty = k\mathbf{A}_{ij}^\infty \mathbf{p}_i^\infty \tag{38}$$

for some $k \neq 0$. Substituting Equation (37) into Equation (38) we obtain the following linear system

$$\mathbf{p}_j^\infty = \left[X\mathbf{A}_{ij}^\star(\pi) + Y \frac{\bar{\mathbf{e}}_{ji}\bar{\mathbf{e}}_{ji}^T}{||\bar{\mathbf{e}}_{ji}||^2} \left(\mathbf{A}_{ij}^\star(\pi) - \mathbf{A}_{ij}^\star(\omega)\right) \right] \mathbf{p}_i^\infty, \tag{39}$$

where $X = k\frac{z_j^0}{z_i^0}$ and $Y = k\frac{z_j^0}{z_i^0}\frac{1}{\alpha}$ are the unknowns. The infinite homography can be estimated up to the scale factor k by solving the system, and k can finally be retrieved by imposing that $\det(\mathbf{A}_{ij}^\infty) = 1$.

Notice that it is important for the vanishing point not to be on the planes. In fact, if $\mathbf{p}_i^\infty \in \boldsymbol{\pi}$ or $\mathbf{p}_i^\infty \in \boldsymbol{\omega}$, it follows that $\mathbf{v}_i^T \mathbf{p}_i^\infty = 0$. Consequently the term that multiplies Y in Equation (39) vanishes, since it is proportional to $\mathbf{v}_i$.

5.2.2. Infinite homography from two pairs of parallel planes

When a further pair of parallel planes $(\boldsymbol{\pi}', \boldsymbol{\omega}')$ is available on the two images, we no more need a vanishing point for estimating the infinite homography. Equation (37) can be rewritten for the second pair of planes, i.e.

$$\mathbf{A}_{ij}^\infty = \frac{z_j^0}{z_i^0} \left[\mathbf{A}_{ij}^\star(\pi') + \frac{1}{\alpha'} \frac{\bar{\mathbf{e}}_{ji}\bar{\mathbf{e}}_{ji}^T}{||\bar{\mathbf{e}}_{ji}||^2} \left(\mathbf{A}_{ij}^\star(\pi') - \mathbf{A}_{ij}^\star(\omega')\right) \right], \tag{40}$$

where, in general, $\alpha' \neq \alpha$. Noticing that $\mathbf{e}_{ji} = \lambda \mathbf{A}_{ij}^\infty \mathbf{e}_{ij}$ for some unknown $\lambda \neq 0$, right-multiplying eqs.(37) and (40) by $\mathbf{e}_{ij}$ leads to the system

$$\begin{cases} \mathbf{e}_{ji} = \lambda \frac{z_j^0}{z_i^0} \left[\mathbf{A}_{ij}^\star(\pi) + \frac{1}{\alpha} \frac{\bar{\mathbf{e}}_{ji}\bar{\mathbf{e}}_{ji}^T}{||\bar{\mathbf{e}}_{ji}||^2} \left(\mathbf{A}_{ij}^\star(\pi) - \mathbf{A}_{ij}^\star(\omega)\right) \right] \mathbf{e}_{ij} \\ \mathbf{e}_{ji} = \lambda \frac{z_j^0}{z_i^0} \left[\mathbf{A}_{ij}^\star(\pi') + \frac{1}{\alpha'} \frac{\bar{\mathbf{e}}_{ji}\bar{\mathbf{e}}_{ji}^T}{||\bar{\mathbf{e}}_{ji}||^2} \left(\mathbf{A}_{ij}^\star(\pi') - \mathbf{A}_{ij}^\star(\omega')\right) \right] \mathbf{e}_{ij} \end{cases}. \tag{41}$$

Notice also that $c_{(\cdot)}\mathbf{e}_{ji} = \mathbf{A}_{ij}^\star(\cdot)\mathbf{e}_{ij}$ for any choice of the reference plane, where the scaling constant $c_{(\cdot)}$ is known if the epipoles and the homography are known. In the light of these considerations, the system can be rewritten as

$$\begin{cases} \mathbf{e}_{ji} &= \lambda \frac{z_j^0}{z_i^0} \left[c_\pi + \frac{1}{\alpha}(c_\pi - c_\omega) \right] \mathbf{e}_{ji} \\ \mathbf{e}_{ji} &= \lambda \frac{z_j^0}{z_i^0} \left[c'_\pi + \frac{1}{\alpha'}(c'_\pi - c'_\omega) \right] \mathbf{e}_{ji} \end{cases},$$

where c_π, c_ω, c'_π and c'_ω are the known scales of $\mathbf{e}_{ji}$ (as an example, $c_\pi \mathbf{e}_{ji} = \mathbf{A}_{ij}^\star(\pi)\mathbf{e}_{ij}$). At this point it is immediate to see that

$$\lambda \frac{z_j^0}{z_i^0} \left[c_\pi + \frac{1}{\alpha}(c_\pi - c_\omega) \right] = \lambda \frac{z_j^0}{z_i^0} \left[c'_\pi + \frac{1}{\alpha'}(c'_\pi - c'_\omega) \right] = 1,$$

and therefore

$$\frac{1}{\alpha'} = \frac{c_\pi - c'_\pi + \frac{1}{\alpha}(c_\pi - c_\omega)}{c'_\pi - c'_\omega}. \tag{42}$$

From Equation (41) it is readily verified that

$$\mathbf{A}_{ij}^\star(\pi) + \frac{1}{\alpha} \frac{\bar{\mathbf{e}}_{ji}\bar{\mathbf{e}}_{ji}^T}{||\bar{\mathbf{e}}_{ji}||^2} \left(\mathbf{A}_{ij}^\star(\pi) - \mathbf{A}_{ij}^\star(\omega)\right) =$$
$$= \mathbf{A}_{ij}^\star(\pi') + \frac{1}{\alpha'} \frac{\bar{\mathbf{e}}_{ji}\bar{\mathbf{e}}_{ji}^T}{||\bar{\mathbf{e}}_{ji}||^2} \left(\mathbf{A}_{ij}^\star(\pi') - \mathbf{A}_{ij}^\star(\omega')\right),$$

from which we can solve for $\frac{1}{\alpha}$ by inserting the expression of $\frac{1}{\alpha'}$ given by Equation (42). The infinite homography can be finally obtained from Equation (37) calculating the ratio $\frac{z_j^0}{z_i^0}$ by posing $\det(\mathbf{A}_{ij}^\infty) = 1$.

5.3. Summary of the view synthesis algorithm

For the sake of clarity, in this section we summarize the presented view synthesis algorithm, starting from the one proposed by Shashua and Navab reported in Section 3. We identify the following steps:

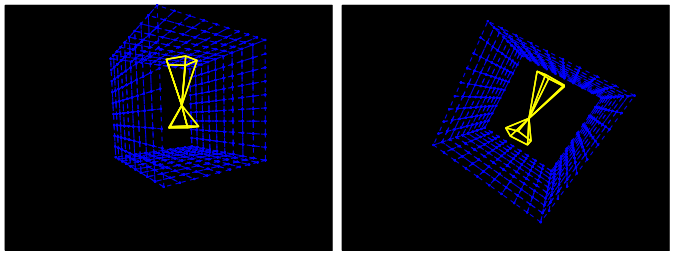
1. From a set of correspondences in the two views, estimate the fundamental matrix and the epipoles $\mathbf{e}_{ij}$ and $\mathbf{e}_{ji}$ with (10);
2. In the original approach [19] defining a reference plane, $\boldsymbol{\pi}$ it is possible to recover the relative affine structure (eq. 11) and then reconstruct the 3D geometry (12). While in the proposed work we follow the approach depicted in the following steps.
3. Due to the belonging to the Special Projectivity Group $SP(3)$ of the transformation $\mathbf{T}$ between two cameras with the same Intrinsic parameters, we can map such a transformation as a Lie Group Homomorphism (eq. 19) and a continuous set of intermediate transformations can be generated.
4. Estimate the infinite homography with the method described either in Sec. 5.2.1 (based on two parallel planes and a vanishing point) or 5.2.2 (based on two pairs of parallel planes) and scale it such that $\det(\mathbf{A}_{ij}^\infty) = 1$;
5. select a reference point correspondence $(\mathbf{x}_i^0, \mathbf{x}_j^0)$ and scale the epipole such that $\mathbf{x}_j^0 \cong \mathbf{A}_{ij}^\infty \mathbf{x}_i^0 + \mathbf{e}_{ji}$;
6. recover the projective reconstruction $\mathbf{X}_i^P$ according to Equation (12);
7. Once the relative transformation matrix $\mathbf{D}_{ij}$ is obtained from one of the proposed methods, it can be made homomorphic to $\mathbf{T}_{ij}$ through the pre-multiplication with the *scene calibration matrix* $\mathcal{V}_j$ that makes their product similar to $\mathbf{T}_{ij}$ (eq. 28).
8. generate the virtual views through Equation (29), namely: $\mathbf{x}(t) \cong [\mathbf{I}|\mathbf{0}](\mathbf{V}_j\mathbf{D}_{ij})^t \mathbf{X}_i^P$.

6. Experiments

This last section is devoted to showing the effectiveness and the robustness of the proposed solutions. First we present a series of experiments based on a 3D synthetic scene, then we test the effectiveness of our method on real images.

6.1. Synthetic setup

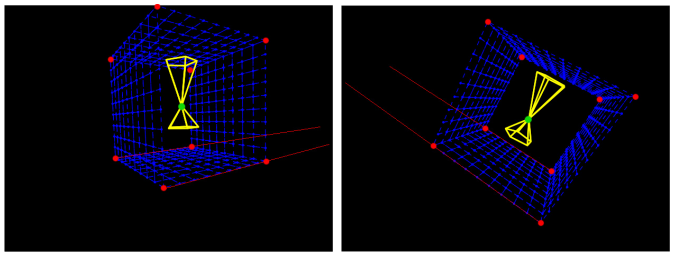
We present the first set of experiments, that are based on a 3D synthetic scene which is projected in the two reference views shown in Figure 4. The reference views have a dimension of 1600×1200 pixels, and are obtained specifying two different camera matrices having the same internal calibration parameters.



(a) View i (b) View j

Figure 4: The two reference views are obtained by projecting a 3D synthetic scene by means of 2 different camera matrices with the same internal calibration parameters.

The synthetic 3D scene is composed by a complex object defined by 11 points (pictured in yellow in Figure 4) enclosed in a cube. Four of the six sides of the cube exhibit a regular grid structure, yielding a total of 411 point correspondences between the two views. This structure allows to identify, in the two images, all the geometric elements that are necessary for estimating the infinite homography. All the simulations have been conducted according to Figure 5: two pairs of parallel planes are defined by the eight red points corresponding to the vertexes of the cube; and a vanishing point is obtained by intersecting the two red lines passing through the bottom vertexes. The green point is selected as a reference for the evaluations.



(a) View i (b) View j

Figure 5: The cube structure allows to identify two pairs of parallel planes and a vanishing point in the two views.

6.2. Evaluation metric

The robustness of the proposed view synthesis algorithm is tested by corrupting all the 411 point correspondences with zero-mean additive Gaussian noise with standard deviation σ, i.e.

$$\tilde{\mathbf{x}}_i^{(k)} = \mathbf{x}_i^{(k)} + \begin{bmatrix} e_x \\ e_y \\ 0 \end{bmatrix} ,$$

where $e_x \sim N(0, \sigma^2)$ and $e_y \sim N(0, \sigma^2)$ are two independent random variables. The noisy data $\tilde{\mathbf{x}}_j$ for the second view are obtained analogously. From the noisy correspondences one obtains the estimate of the camera calibration matrix $\mathbf{V}_i$ and of the transformation $\mathbf{D}_{ij}$, namely the matrices $\tilde{\mathbf{V}}_j$ and $\tilde{\mathbf{D}}_{ij}$, respectively. The view synthesis algorithm is evaluated by considering the error between the exact position and the reconstructed position of the reference points in the virtual view. The error is calculated by means of the Euclidean distance between the target coordinates $\mathbf{x}(t) = [x(t), y(t), 1]^T$ obtained exponentiating the matrix $\mathbf{T}_{ij}$ by means of Equation (22) and the reconstruction $\hat{\mathbf{x}}(t) = [\hat{x}(t), \hat{y}(t), 1)]^T$ through the matrix $\tilde{\mathbf{V}}_j \tilde{\mathbf{D}}_{ij}$ by means of Equation (29), i.e.:

$$d_{t,\sigma} = \sqrt{[x(t) - \hat{x}(t)]^2 + [y(t) - \hat{y}(t)]^2} \, . \qquad (43)$$

The evaluation is performed with different noise levels (σ from 0 to 4 pixels) and considering virtual views generated by varying the parameter t in the range from 0 to 1. The performances of the algorithm are finally summarized calculating the mean error and the root mean squared error (RMSE) applied to the distance measure defined in Equation (43), averaging the results over 1000 realizations.

6.3. Results

In this section we show view synthesis results relative to the two methods for estimating the infinite homography.

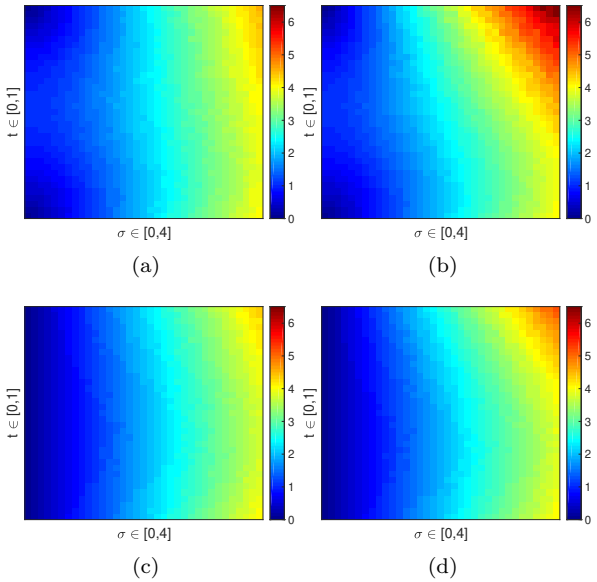


Figure 6: Performances of the view synthesis algorithm relative to the RMS reconstruction error of the reference point. Top row (a,b): The infinite homography is estimated from a pair of parallel planes and a vanishing point. Bottom row (c,d): The infinite homography is estimated from two pairs of parallel planes. Left column: the matrix $\mathbf{D}_{ij}$ is calculated using Equation (24). Right column: the matrix $\mathbf{D}_{ij}$ is estimated from point correspondences.

Consider first the estimation by means of two parallel planes and a vanishing point. The results are drawn

both in the case that the matrix $\mathbf{D}_{ij}$ is estimated through Equation (24) and for the estimation of $\mathbf{D}_{ij}$ from the set of correspondences $(\mathbf{X}_i^P, \mathbf{X}_j^P)$. The correspondent results are shown in Figure 6.

It can be observed that the view synthesis is very robust against the noise on the point correspondences, especially when the matrix $\mathbf{D}_{ij}$ is estimated through its expression (Figure 6(a)): in this case, when the error is set to $\sigma = 4$ pixels (right side of the image), the RMSE maintains lower than 5 pixels. Notice that, as expected, the reconstruction error is generally lower for the synthetic views close to the reference ($t = 0$, bottom of the image), and it tends to increase for views with increasing t. Regarding the second method for estimating $\mathbf{D}_{ij}$, Figure 6(b) reveals that, in this case, the performances are slightly worse.

Consider now the case of estimating the infinite homography from two pairs of parallel planes. As before, the results are presented for the two methods for estimating the matrix $\mathbf{D}_{ij}$, and are reported in Figures 6(c) and 6(d). When $\mathbf{D}_{ij}$ is explicitly calculated, the algorithm exhibits less sensibility to the noise, especially for high values of σ. However, when $\sigma < 2$ pixels, the results are comparable to the case of Figures 6(a) and 6(b). When the matrix $\mathbf{D}_{ij}$ is estimated from the correspondences, we observe that the results revert to be comparable to the case of a pair of parallel planes and a vanishing point.

6.4. Comparison of the proposed approaches

In this section the two proposed methods are compared more closely. For these experiments we set the parameter $t = 0.5$ and $\mathbf{D}_{ij}$ is estimated from the correspondences (as in Figures 6(b) and 6(d)), and we use all of the 11 points of the yellow object as reference points.

We included a strategy where the infinite homography is computed using the perspective projection matrices (PPMs) of the two views. Let $P_i := [Q_i|q_i]$ where $i = 1, 2$, be the PPMs, then the infinite homography between the two reference views can be calculated according to the formula $\mathbf{A}_{12}^{\infty} := Q_2 Q_1^{-1}$ (see [16] for a full derivation). This equation represents the optimal strategy to compute the infinite homography but it requires the knowledge of fully calibrated PPMs. The reason for its optimality is that it does not involve any information that needs to be estimated from the images, hence any reconstruction error is to be attributed to computation of the relative affine structure.

The first comparison is the noise tolerance of the three methods (the optimal, and the two proposed approaches). We gradually added noise to the point correspondences, as described in Section 6.2, and measured the RMSE of the reconstructed points. Figure 7(a) shows the curve obtained, which shows that the three methods are equally tolerant to correspondence noise.

In the second experiment, instead of corrupting the correspondences coordinates, we introduced noise during the generation of the points that are supposed to lie on the same plane. Please note that in this case the noise is injected in the 3D space, hence the standard deviation σ is not measured in pixels. We used an increasing σ from 0 to 0.2 units of the 3D space. The same noise is applied to the vertices that are used for the computation of the vanishing points. Notice that this kind of corruption should not influence the optimal strategy since it leaves the correspondences precision unaltered.

In Figure 7(b) we can observe that both the proposed methods deteriorate their RMSE score as the noise level applied to the planes increases, but the method that uses one pair of parallel planes and a vanishing point is considerably less resilient than the one that uses two sets of parallel planes. The optimal method, as expected, is insensitive to this kind of corruption.

6.5. View synthesis examples

In this section we show some examples of virtual views generated starting from the reference images of Figure 4. In order to provide a ground-truth, Figure 9(a) reports a set of 9 views obtained exponentiating the theoretical transformation matrix $\mathbf{T}_{ij}$ and applying it to the real 3D model of the scene; the parameter t ranges from 0 to 1. Notice that for $t = 0$ and $t = 1$ the synthesized views corresponds to the references depicted in Figure 4.

Figures 9(b), 9(c) and 9(d) show the 9 novel views generated with the view synthesis algorithm for 3 different choices of the noise level ($\sigma = 1, 2, 3$ respectively). All the synthesized views are obtained estimating $\mathbf{D}_{ij}$ from its mathematical expression, and the infinite homography is retrieved from two pairs of parallel planes.

All the synthetic views confirm the effectiveness of the algorithm. Even for very noisy data ($\sigma = 3$), the reconstructions are not far from the ground-truth, confirming the robustness of the view synthesis algorithm as well as of the methods for estimating the infinite homography. We recall that the aim of the methods presented in this paper is to estimate the interpolated position of corresponding points matched between the two original views; therefore the full synthesis of an interpolated image is not covered.

6.6. Evaluation on real images

In this section we evaluate our approach on a more realistic scenario, where we rely on the presence of two sets of parallel planes. The images were captured using a hand-held Canon EOS 1100D camera. We processed five sequences: "Milo1" is a *ad hoc* assembled scene with a toy figure and a textured cube. The remaining four are hallway scenes, two from the University of Verona and two from the University of Udine[4].

For the sake of comparison, we created an output that is equivalent to our view synthesis using a calibrated approach. In particular, we fed the pictures to a state of the

[4]Please contact the authors to receive a copy of the full resolution images and/or Matlab code.

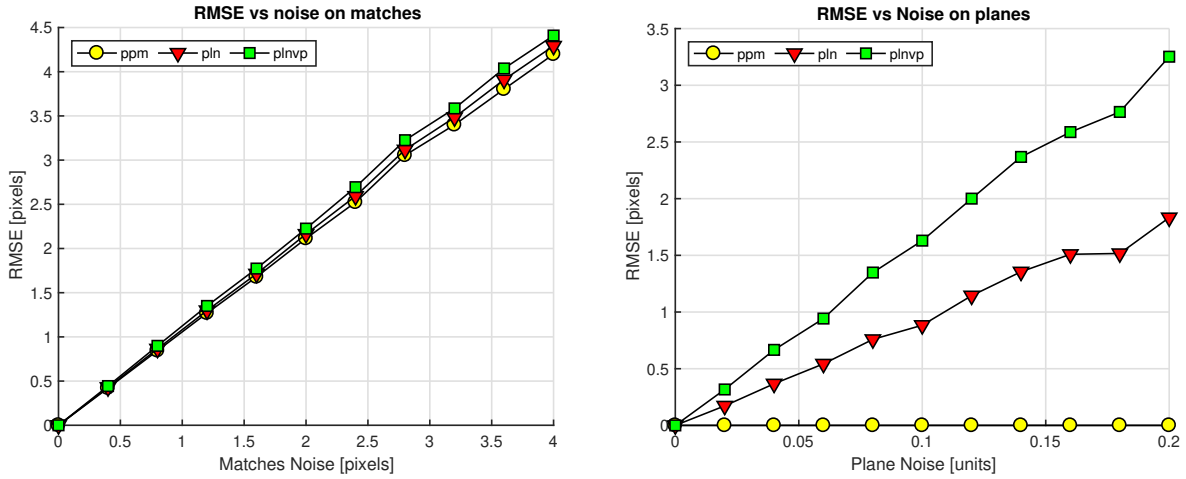


Figure 7: Performances of the view synthesis algorithm relative to the reconstruction error of the reference points. "gt" represent the optimal strategy, "plnvp" is the method that uses a vanishing point and pair of parallel planes and "pln" is the method that uses two pairs of parallel planes.

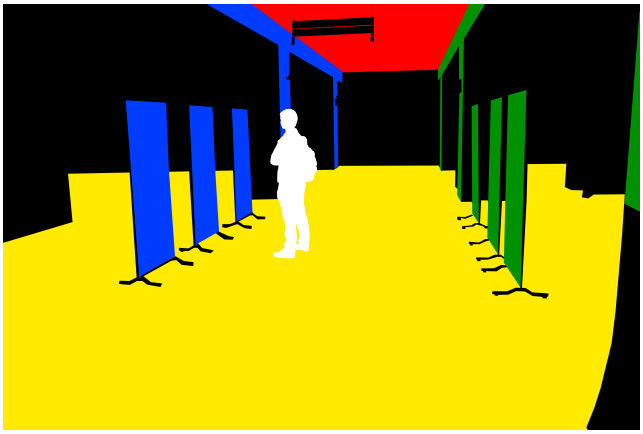


Figure 8: Each color corresponds to a label which is used to identify the 3D points that belong to one of the four planes or to other objects

art structure from motion pipeline ([23], software courtesy of 3Dflow srl). For each scene, we acquired several images with known, constant internal parameters, from which we obtained a sparse 3D reconstruction of the scene, i.e. a set of 3D points. Then, we proceeded by manually segmenting the points lying on the two sets of parallel planes in the 3D space (see Figure 8).

Such manual selection was then validated through a RANSAC procedure that fitted the dominant planar homography, and was used to discard points that were not pertinent with the plane. On the left side of Figure 10 one can appreciate the output of this step. Then, among the initial batch, we selected two images to be used as the two reference views for the view synthesis experiment. The keypoints corresponding to the 3D points that are visible in the two images are used as reference for the performance evaluation. The reference keypoints are transferred to the virtual view point using Equation (15) (internal parameters are known only for the sake of evaluation) and setting $t = 0.5$. The central column of Figure 10 shows the reference images of each sequence with a subset of the corresponding keypoints. Finally, the points obtained through the view synthesis procedure are then compared against reference points to obtain the RMSE, as in the synthetic experiments. The effectiveness of the proposed approach can be appreciated by observing the right side of Figure 10 that shows a visual comparison of the positioning of the keypoints on the virtual view with respect to the ground truth points. In Table 1 we report the RMSE scores for each sequence.

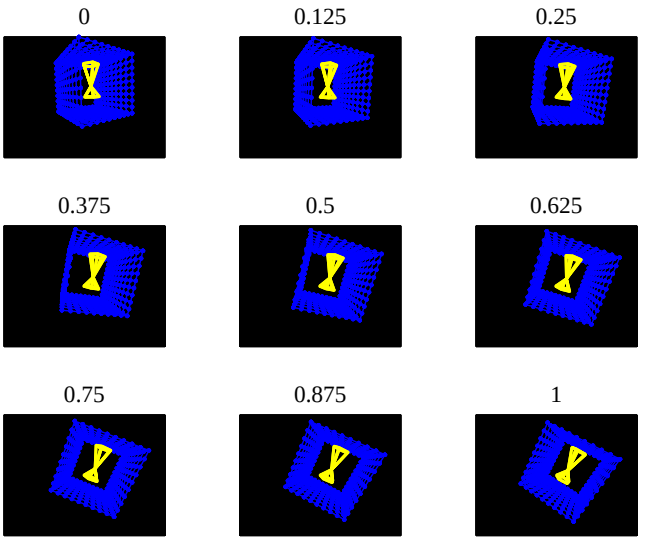
Table 1: RMSE scores obtained for the five proposed sequences.

Sequence	# points	t	RMSE
UniVR1	6942	0.5	0.0184
UniVR2	20418	0.5	0.3664
UniUD1	2594	0.5	0.2326
UniUD2	3236	0.5	0.6409
Milo1	28334	0.5	0.0278

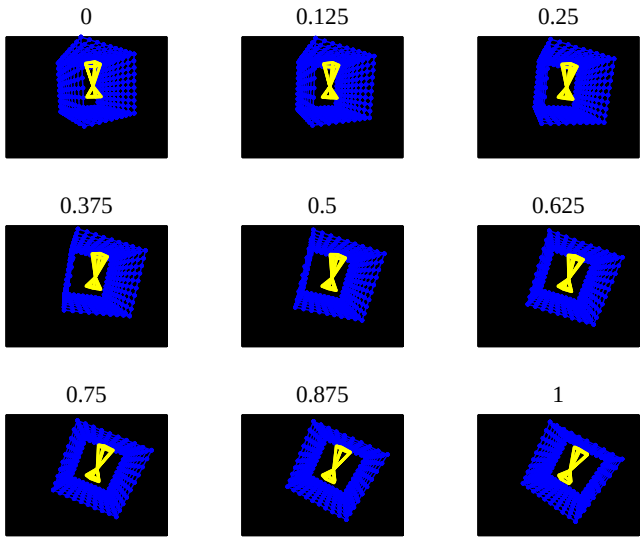
7. Conclusions

In this work we provide a full discussion on the generation of synthetic images from uncalibrated cameras. The proposed algorithm allows to smoothly interpolate between two reference views provided that the two cameras have identical internal calibration.

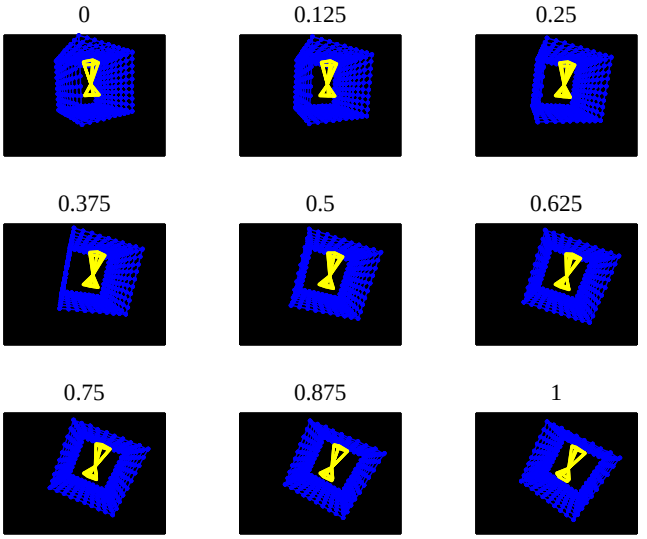
The core contribution of this paper is the proof that the knowledge of infinite homography is sufficient to synthesize view from two uncalibrated images, thus eliminating the need to specify a virtual camera in the projective frame or the need to exploit further geometric constraints. This contribution is substantiated by two novel methods



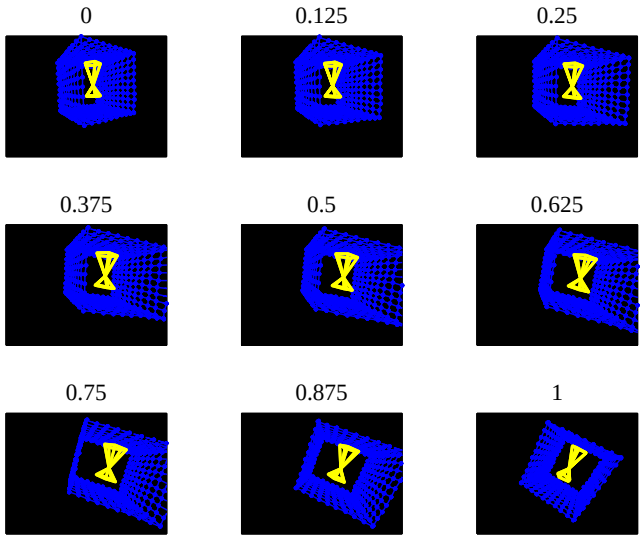
(a) Ground truth.



(b) $\sigma = 1$ pixels.



(c) $\sigma = 2$ pixels.



(d) $\sigma = 3$ pixels.

Figure 9: Virtual views obtained exponentiating the theoretical transformation matrix $\mathbf{T}_{ij}$ and applying it to the real 3D model of the scene. The parameter t ranges from 0 to 1. This set of 9 views constitute a ground-truth for evaluating the view synthesis algorithm. Virtual views obtained with the view synthesis algorithm. The infinite homography is estimated from two pairs of parallel planes. The noise level is set to $\sigma = 1, 2, 3$ pixel.

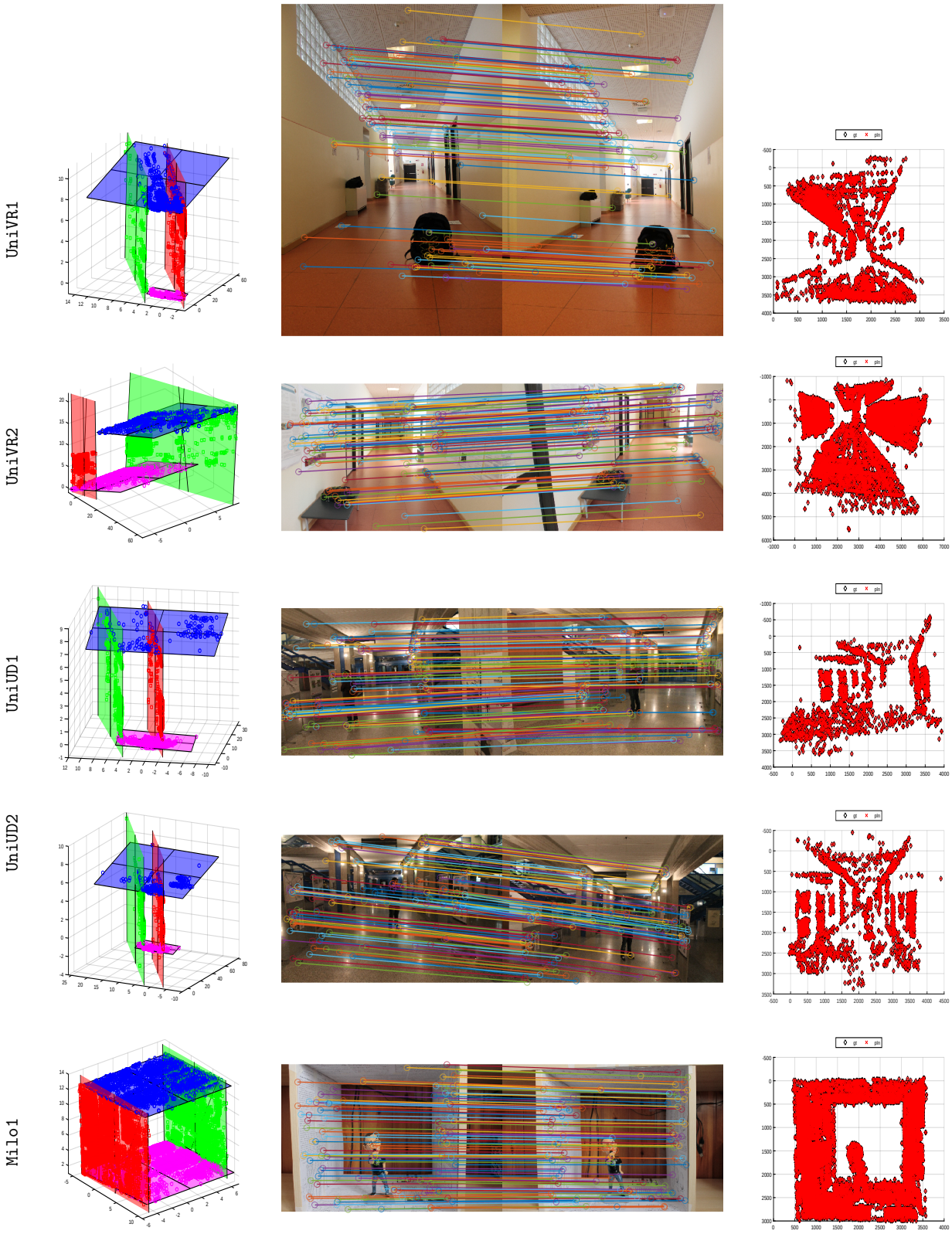


Figure 10: Left column: the result of the plane segmentation step. Central column: reference images with matched keypoints. Right column: comparison of the points transferred with the view synthesis procedure (red crosses) against the ground truth points (black diamonds).

for computing infinite homography, with prior information of (i) two pairs of parallel planes and (ii) one pair of parallel planes plus vanishing points not on the planes.

Depending on the geometric elements present in the imaged scene, one can select the best technique for estimating the infinite homography.

The simulations conducted on a synthetic 3D scene and the experiments on real images confirm the effectiveness and the robustness of the proposed approach.

References

[1] M. Alexa. Linear combination of transformations. *Proc. 29th Ann. Conf. Comp. Graph. Interactive Techn.*, pages 380–387, 2002.

[2] S. Avidan and A. Shashua. Novel view synthesis in tensor space. *Conference on computer vision and pattern recognition*, pages 1034–1040, June 1997.

[3] S. Avidan and A. Shashua. Novel view synthesis by cascading trilinear tensors. *IEEE Trans. on Visualization and Computer Graphics*, 4:293–306, Oct. 1998.

[4] A. Colombari, A. Fusiello, and V. Murino. Uncalibrated interpolation of rigid displacements for view synthesis. *Proc. IEEE Int. Conf. Image Process. (ICIP 2005)*, pages 1049–1052, 2005.

[5] K. Connor and I. Reid. Novel view specification and synthesis. In *British Machine Vision Conference*, pages 243–252, 2002.

[6] R. B. Fisher. The camera project (cad modelling of built environments from range analysis). *http://www.dai.ed.ac.uk/homes/rbf/CAMERA/camera.htm.*

[7] H. Fuchs, G. Bishop, K. Arthur, L. McMillan, R. Bajcsy, S. Lee, H. Farid, and T. Kanade. Virtual space teleconferencing using a sea of cameras. *First International Symposium on Medical Robotics and Computer Assisted Surgery*, pages 161–167, 1994.

[8] A. Fusiello. Specifying virtual cameras in uncalibrated view synthesis. *IEEE Trans. on Circuit and Systems for Video Technology*, 17:604–611, May 2007.

[9] J. Gallier. Manifolds, lie groups, lie algebras, riemannian manifolds, with applications to computer vision and robotics. *http://www.cis.upenn.edu/ cis610*, 2005.

[10] R. Gilmore. Lie groups: general theory. *Encyclopedia of Mathematical Physics*, EMP MS 425, Feb. 2005.

[11] D. Hogg. The resolv project (reconstruction using scanned laser and video). *http://www.scs.leeds.ac.uk/resolv.*

[12] Y. J. Jeong, H. Hwang, D. Nam, and C. . J. Kuo. Uncalibrated multiview synthesis based on epipolar geometry approximation. In *2015 IEEE International Conference on Consumer Electronics (ICCE)*, pages 542–543, Jan 2015.

[13] J.-S. Kim and I. Kweon. Infinite homography estimation using two arbitrary planar rectangles. volume 3852, pages 1–10, 2006.

[14] M. Levoy. The digital michelangelo project: 3d scanning of large statues. *http://graphics.stanford.edu/projects/mich.*

[15] L. McMillan and G. Bishop. Plenoptic modeling: An image-based rendering system. *SIGGRAPH 95 Conference Proceedings*, pages 39–46, Aug. 1995.

[16] A. Z. R. Hartley. Multiple view geometry in computer vision. *Cambridge University Press*, 2000.

[17] S. Seitz and C. Dyer. Phisically-valid view synthesis by image interpolation. *In Workshop on Representations of Visual Scenes*, 1995.

[18] S. M. Seitz and C. R. Dyer. View morphing: Synthesizing 3D metamorphoses using image transforms. In *SIGGRAPH: International Conference on Computer Graphics and Interactive Techniques*, pages 21–30, New Orleans, Louisiana, Aug. 1996.

[19] A. Shashua and N. Navab. Relative affine structure: Canonical model for 3d from 2d geometry and applications. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 18:873–883, Sept. 1996.

[20] R. S. S.J. Gortler, R. Grzeszczuk and M. Choen. The lumigraph. *Computer Graphics (SIGGRAPH'96)*, pages 43–54, Aug. 1996.

[21] P. R. T. Kanade, P. Narayanan. Virtualized reality: Concepts and early results. *In IEEE Workshop on Representation of Visual Scenes*, pages 69–76, June 1995.

[22] S. Tebaldini, M. Marcon, A. Sarti, and S. Tubaro. Uncalibrated view synthesis from relative affine structure based on planes parallelism. In *Image Processing, 2008. ICIP 2008. 15th IEEE International Conference on*, pages 317 –320, oct. 2008.

[23] R. Toldo, R. Gherardi, M. Farenzena, and A. Fusiello. Hierarchical structure-and-motion recovery from uncalibrated images. *Computer Vision and Image Understanding*, 140:127 – 143, November 2015.

[24] Y. Zhao, S. Wang, J. Wang, and H. Ding. Linear solving the infinite homography matrix from epipole. volume 1, pages 92 –96, jan. 2010.