VIEW SYNTHESIS ALONG A CURVE FROM TWO UNCALIBRATED VIEWS

Andrea Fusiello and Andrea Colombari

Dipartimento di Informatica, Università di Verona
Strada Le Grazie, 15 - 37134 Verona, Italy
andrea.fusiello@univr.it, colombari@sci.univr.it

ABSTRACT

This paper presents a generic framework for novel view synthesis from two uncalibrated reference views that allows to move a virtual camera along a curve that is obtained starting from the epipolar geometry of the reference views. The scene is described by its relative affine structure from which novel views are extrapolated and interpolated. The main contribution of this paper is an automatic method for specifying virtual camera locations in an uncalibrated setting. Experiments with synthetic and real images illustrate the approach.

1. INTRODUCTION

Nowadays, we witness an increasing interest in the convergence of Computer Vision and Computer Graphics, and, in this stream, one of the most promising and fruitful area is *Image-Based Rendering* (IBR) [1]. While the traditional geometry-based rendering starts from a 3-D model, in IBR views are generated by re-sampling one or more example images, using appropriate warping functions [2].

In the case of calibrated cameras, algorithms based on image interpolation yield satisfactory results [3, 4]. Where no knowledge on the imaging device can be assumed, uncalibrated point transfer techniques utilize image-to-image constraints such as the Fundamental matrix [5], trilinear tensors [6], plane+parallax [7], to reproject pixels from a small number of reference images to a given view. Another way of linking corresponding points is the *relative affine structure* [8], a close relative of the plane+parallax. This is the framework in which our technique is embedded.

Although uncalibrated point transfer algorithms are well understood, what prevent them to be applied in real-world applications, is the lack of a "natural" way of specifying the position of the virtual camera in the familiar Euclidean frame, because it is not accessible. Everything is represented in a projective frame that is linked to the Euclidean one by an *unknown* projective transformation. All the view-synthesis algorithms requires either to manually input the position of points in the synthetic view, or to specify some projective elements.

In this work, we will consider the case of interpolation and extrapolation from two uncalibrated reference views. We propose a solution to the specification of the new viewpoints, based on the exploitation of the epipolar geometry that links the reference views, represented by the homography of the plane at infinity and the epipole. Thanks to the group structure of these *uncalibrated rigid transformations*, interpolation and extrapolation is possible using matrix exponential and logarithm. The virtual cameras are positioned as if the real camera continued with the same motion as between the two reference views.

Our technique allows to synthesize physically-valid views, and in this sense it can be seen as a generalization to the uncalibrated case of [4]. The framework for interpolation of Euclidean transformations was set forth in [9], whereas the idea of manipulating rigid displacements at the uncalibrated level is outlined in [10], where it is applied to rotations only.

This work is particularly significant in the context of stereoscopic visualization, like in 3-D television, where two separate video streams are produced, one for each eye. In order to avoid viewer's discomfort, the amount of parallax encoded in the stereo pair must be adapted to the viewing condition, or, equivalently, the virtual viewpoint needs to be moved. The idea is that the viewer might use a "3D-ness" knob [11] to continuously adjust the stereoscopic separation. Uncalibrated view-synthesis offers a solution that does not require the reconstruction of the full scene structure, but only the estimation of disparities.

This work has been supported by the LIMA3D project (Progetto di Ricerca di Interesse Nazionale (PRIN) 2003.

2. BACKGROUND

We start by giving some background notions needed to understand our method. A complete discussion on the relative affine structure theory can be found in [8].

Given a plane Π, with equation $\boldsymbol{n}^{\mathsf{T}}\boldsymbol{w} = d$, two conjugate points $\boldsymbol{m}_1$ and $\boldsymbol{m}_2$ are related by

$$\boldsymbol{m}_2 \backsim \boldsymbol{H}_{12}\boldsymbol{m}_1 + \boldsymbol{e}_{21}\gamma \tag{1}$$

where $\boldsymbol{H}_{12}$ is the collineation induced by a plane Π and $\boldsymbol{e}_{21}$ is the the epipole in the second view. The symbol $\backsim$ means equality up to a scale factor. If the 3D point $\boldsymbol{w} \notin \Pi$, there is a residual displacement, called *parallax*. This quantity is proportional to the *relative affine structure* $\gamma \triangleq \dfrac{a}{d\,\kappa_1}$ of $\boldsymbol{w}$ [8], where a is the orthogonal distance of the 3-D point $\boldsymbol{w}$ to the plane Π and κ_1 is the distance of $\boldsymbol{w}$ from the focal plane of the first camera. Points $\boldsymbol{m}_2$, $\boldsymbol{H}_{12}\boldsymbol{m}_1$ and $\boldsymbol{e}_{21}$ are collinear. The parallax field is a radial field centered on the epipole.

Since the relative affine structure is independent on the second camera, arbitrary "second views" can be synthesized, by giving a plane homography and an epipole, which specify the position and orientation of the virtual camera in a projective framework. The view synthesis algorithm that we employ, inspired by [8], is the following:

A. given a set of conjugate pairs $(\boldsymbol{m}_1^\ell\,; \boldsymbol{m}_2^\ell) \quad \ell = 1, \ldots, m$;
B. recover the epipole $\boldsymbol{e}_{21}$ and the homography $\boldsymbol{H}_{12}$ up to a scale factor;

C. choose a point $\boldsymbol{m}_1^0$ and scale $\boldsymbol{H}_{12}$ to satisfy

$$\boldsymbol{m}_2^0 \backsim \boldsymbol{H}_{12}\boldsymbol{m}_1^0 + \boldsymbol{e}_{21}$$

D. compute the relative affine structure γ^ℓ from (1):

$$\gamma^\ell = \frac{(\boldsymbol{m}_2^\ell \times \boldsymbol{e}_{21})^T(\boldsymbol{H}_{12}\boldsymbol{m}_1^\ell \times \boldsymbol{m}_2^\ell)}{||\boldsymbol{m}_2^\ell \times \boldsymbol{e}_{21}||^2}. \qquad (2)$$

E. specify a new epipole $\boldsymbol{e}_{31}$ and a new homography $\boldsymbol{H}_{13}$ (properly scaled);

F. transfer points in the synthetic view with

$$\boldsymbol{m}_3^\ell \backsim \boldsymbol{H}_{13}\boldsymbol{m}_1^\ell + \boldsymbol{e}_{31}\gamma^\ell \qquad (3)$$

The problem that makes this technique difficult to use in practice (and for this reason it has been overlooked for view synthesis) is point E, namely that one has to specify a new epipole $\boldsymbol{e}_{31}$ and a new (scaled) homography $\boldsymbol{H}_{13}$. In Section 3 we will present an automatic solution to this problem.

3. SPECIFYING THE VIRTUAL CAMERA POSITION

Our idea is based on the replication of the unknown rigid displacement $\boldsymbol{G}_{12}$ that links the reference views, I_1 and I_2. The method described in this section will allow us to render a view I_3 from a pose $\boldsymbol{G}_{13} = \boldsymbol{G}_{12}\boldsymbol{G}_{12} = (\boldsymbol{G}_{12})^2$. More in general, thanks to the group structure, this will extends to to any scalar multiple of $\boldsymbol{G}_{12}$,

3.1. The group of uncalibrated rigid displacements

Let us consider Eq. (1), which express the epipolar geometry with reference to a plane, in the case of view pair 1-2:

$$\frac{\kappa_2}{\kappa_1}\boldsymbol{m}_2 = \boldsymbol{H}_{12}\boldsymbol{m}_1 + \boldsymbol{e}_{21}\gamma_1 \qquad (4)$$

and view pair 2-3:

$$\frac{\kappa_3}{\kappa_2}\boldsymbol{m}_3 = \boldsymbol{H}_{23}\boldsymbol{m}_2 + \boldsymbol{e}_{32}\gamma_2. \qquad (5)$$

In order to obtain an equation relating view 1 and 3, let us substitute the first into the second, obtaining:

$$\frac{\kappa_3}{\kappa_1}\boldsymbol{m}_3 = \boldsymbol{H}_{23}\boldsymbol{H}_{12}\boldsymbol{m}_1 + (\boldsymbol{H}_{23}\boldsymbol{e}_{21} + \boldsymbol{e}_{32}\frac{d_1}{d_2})\gamma_1 \qquad (6)$$

By comparing this equation to Eq. (1), we obtain:

$$\boldsymbol{e}_{31} = \boldsymbol{H}_{23}\boldsymbol{e}_{21} + \boldsymbol{e}_{32}\frac{d_1}{d_2} \qquad (7)$$

The ratio $\frac{d_1}{d_2}$ in general is unknown, but if Π is the plane at infinity then $\frac{d_1}{d_2} = 1$ (please note that this is approximatly true for planes distant from the camera). Therefore, taking the plane at infinity as Π, from Eq. (6) we obtain:

$$\begin{aligned} \boldsymbol{H}_{\infty 13} &= \boldsymbol{H}_{\infty 23}\boldsymbol{H}_{\infty 12} \\ \boldsymbol{e}_{31} &= \boldsymbol{H}_{\infty 23}\boldsymbol{e}_{21} + \boldsymbol{e}_{32} \end{aligned} \qquad (8)$$

In matrix form Eq. (8) writes:

$$\boldsymbol{D}_{13} = \boldsymbol{D}_{23}\boldsymbol{D}_{12} \qquad (9)$$

where

$$\boldsymbol{D}_{ij} \triangleq \begin{bmatrix} \boldsymbol{H}_{\infty ij} & \boldsymbol{e}_{ji} \\ \boldsymbol{0} & 1 \end{bmatrix} \qquad (10)$$

represents a *rigid displacement at the uncalibrated level*[1].

We then plug $\boldsymbol{D}_{13}$ as defined above in the transfer equation (Eq. (3)) that re-writes:

$$\boldsymbol{m}_3^\ell \backsim \boldsymbol{D}_{13}\begin{bmatrix} \boldsymbol{m}_1^\ell \\ \gamma_1^\ell \end{bmatrix} \qquad (11)$$

We will now prove that the virtual view I_3 obtained from the above equation is rendered from a pose $\boldsymbol{G}_{13} = \boldsymbol{G}_{23}\boldsymbol{G}_{13}$. Let

$$\boldsymbol{G}_{ij} \triangleq \begin{bmatrix} \boldsymbol{R}_{ij} & \boldsymbol{t}_{ij} \\ \boldsymbol{0} & 1 \end{bmatrix} \qquad (12)$$

be a matrix that represents a rigid displacement, where $\boldsymbol{R}$ is a rotation matrix and $\boldsymbol{t}$ is a vector representing a translation. Rigid displacements form a group, known as the special Euclidean group of rigid displacements in 3D, denoted by $SE(3)$.

Each uncalibrated displacement $\boldsymbol{D}_{ij}$ is the conjugate of an element $\boldsymbol{G}_{ij} \in SE(3)$ by the matrix $\tilde{\boldsymbol{A}} = \begin{bmatrix} \boldsymbol{A} & \boldsymbol{0} \\ \boldsymbol{0} & 1 \end{bmatrix}$:

$$\begin{aligned} \boldsymbol{D}_{ij} &= \begin{bmatrix} \boldsymbol{A}\boldsymbol{R}_{ij}\boldsymbol{A}^{-1} & \boldsymbol{A}\boldsymbol{t}_{ij} \\ \boldsymbol{0} & 1 \end{bmatrix} \\ &= \begin{bmatrix} \boldsymbol{A} & \boldsymbol{0} \\ \boldsymbol{0} & 1 \end{bmatrix}\begin{bmatrix} \boldsymbol{R}_{ij} & \boldsymbol{t}_{ij} \\ \boldsymbol{0} & 1 \end{bmatrix}\begin{bmatrix} \boldsymbol{A}^{-1} & \boldsymbol{0} \\ \boldsymbol{0} & 1 \end{bmatrix} \end{aligned}$$

The conjugacy (or similarity) mapping is an homomorphism of $SE(3)$, for it preserves the product:

$$\begin{aligned} \boldsymbol{D}_{13} = \boldsymbol{D}_{23}\boldsymbol{D}_{12} &= \tilde{\boldsymbol{A}}\boldsymbol{G}_{23}\tilde{\boldsymbol{A}}^{-1}\tilde{\boldsymbol{A}}\boldsymbol{G}_{12}\tilde{\boldsymbol{A}}^{-1} \\ &= \tilde{\boldsymbol{A}}\boldsymbol{G}_{23}\boldsymbol{G}_{12}\tilde{\boldsymbol{A}}^{-1} = \tilde{\boldsymbol{A}}\boldsymbol{G}_{13}\tilde{\boldsymbol{A}}^{-1}. \end{aligned} \qquad (13)$$

This proves our thesis and also points out the conjugacy relationship between $SE(3)$ and the group of uncalibrated displacement.

3.2. Extrapolation and interpolation

Let us focus on the problem of specifying the virtual camera's viewpoint. Please note that if intrinsic parameters are constant, the scale factor of $\boldsymbol{H}_{\infty 12}$ is fixed, since $\det(\boldsymbol{H}_{\infty 12}) = 1$ (see [12]). So, point C in the general view synthesis procedure must be replaced with

C. scale $\boldsymbol{H}_{\infty 12}$ such that $\det(\boldsymbol{H}_{\infty 12}) = 1$.

As for point0 E, there are no free scale factors to fix, as formulas defined in (8) hold with the equality sign.

In the case of synthesis from two views, we know only $\boldsymbol{D}_{12}$ and want to specify $\boldsymbol{D}_{13}$ to be used in the transfer equation to synthesize the 3rd view. The replication trick consist in setting $\boldsymbol{D}_{23} = \boldsymbol{D}_{12}$, i.e., $\boldsymbol{D}_{13} = (\boldsymbol{D}_{12})^2$ thereby obtaining a novel view from a virtual camera placed at $(\boldsymbol{G}_{12})^2$ with respect to the first camera. Likewise, $(\boldsymbol{D}_{12})^n\ \forall n \in \mathbb{Z}$ corresponds to the rigid displacement $(\boldsymbol{G}_{12})^n$.

Integer exponents provide us with an extrapolation scheme by discrete steps. However, $SE(3)$ is also a differentiable manifold

[1]Technically, since we assume to know the plane at infinity, this correspond to the affine calibration stratum [12].

(being a Lie group), in which we can make sense of the interpolation between two elements as drawing the geodesic path between them. Let us consider, without loss of generality, the problem of interpolating between the element $\boldsymbol{G}$ and the identity $\boldsymbol{I}$. The geodesic path leaving the identity can be obtained as the projection of a straight path in the tangent space, and the logarithm map precisely projects a neighborhood of $\boldsymbol{I}$ into the tangent space to $SE(3)$ at $\boldsymbol{I}$. A straight path in the tangent space emanating from 0 is mapped onto a geodesic in $SE(3)$ emanating from $\boldsymbol{I}$ by the exponential map. Hence, the geodesic path in $SE(3)$ joining $\boldsymbol{I}$ and $\boldsymbol{G}$ is given by

$$\boldsymbol{G}^t \triangleq \exp(t \log(\boldsymbol{G})), \quad t \in [0, 1]. \tag{14}$$

More in general, we can define a *scalar multiple of rigid transformations* [9]:

$$t \odot \boldsymbol{G} \triangleq \boldsymbol{G}^t = \exp(t \log(\boldsymbol{G})), \quad t \in \mathbb{R}. \tag{15}$$

Mimicking the definition that we have done for rigid transformations, let us define

$$t \odot \boldsymbol{D} \triangleq \boldsymbol{D}^t = \exp(t \log(\boldsymbol{D})), \quad t \in \mathbb{R}. \tag{16}$$

If we use $\boldsymbol{D}_{1i}(t) = t \odot \boldsymbol{D}_{12}$ in the synthesis, as t varies we obtain a continuous path that interpolates between the two real views for $t \in [0, 1]$, and extrapolates the seed displacement for $t > 1$ or $t < 0$. In this way we are able to move the uncalibrated virtual camera continuously on a curve that passes trough both camera centres. The parameter t is the '3D-ness" knob that we mentioned in the Introduction.

At a calibrated level, this is equivalent to moving the camera along the trajectory $t \odot \boldsymbol{G}$. Indeed,

$$\begin{aligned} \boldsymbol{D}^t = (\tilde{\boldsymbol{A}}\boldsymbol{G}\tilde{\boldsymbol{A}}^{-1})^t = e^{t \log(\tilde{\boldsymbol{A}}\boldsymbol{G}\tilde{\boldsymbol{A}}^{-1})} = \\ = \tilde{\boldsymbol{A}} e^{(t \log \boldsymbol{G})} \tilde{\boldsymbol{A}}^{-1} = \tilde{\boldsymbol{A}}\boldsymbol{G}^t \tilde{\boldsymbol{A}}^{-1}. \end{aligned} \tag{17}$$

A very special case is when the reference views are rectified. Given that no rotation between the two cameras is present, the virtual camera can only be translated along the line containing the centres of the cameras (baseline).

4. IMAGE METHODS

Our technique requires that from the two given images one recovers the epipole, the infinity plane homography and the relative affine structure. We give here a sketch of the methods that we used in our experiments, which are better explained in [13]. Any other method(s) that are able to compute these information from the images could be applied.

Assuming that the background area in the images is bigger than the foreground area, the homography of the background plane is the one that explains the *dominant motion*. We are here implicitly assuming that the background is approximately planar, or that its depth variation is much smaller than its average distance from the camera. We also assume that the background is sufficiently far away so that its homography approximates well the homography of the plane at infinity [14].

After aligning the input images with respect to the background plane, the residual parallax allows to segment off-plane points (foreground). From this we are compute the epipoles and the relative affine structure for the (sparse) set of foreground points. The dense relative affine structure for all the points of the foreground is obtained by interpolation.

Then the foreground is warped using the transfer equation and pixel "splatting" [15]. Pixels are transferred in order of increasing parallax, so that points closer to the camera overwrites farther points.

The planar background is warped using the background homography with destination scan and bilinear interpolation. By warping the background of the second view onto the first one, a mosaic representing all the available information about the background plane is built. Since the foreground could occlude a background area in both the input images, the mosaic could contain holes that are filled by *inpainting*.

5. RESULTS

We report here (Figure 1) some examples of interpolation and extrapolation from real uncalibrated stereo pairs. Albeit parallax adjustment required by 3D-TV is not so considerable, this examples are shown to illustrate the geometrical behaviour of the method.

Our technique makes possible to create an entire sequence as taken by a smoothly moving virtual camera, by continuously changing parameter t in Eq. (16), as illustrated by movies available from http://www.sci.univr.it/~fusiello/demo/synth.

6. CONCLUSION

We presented a technique for the specification of novel viewpoints in the generation of synthetic views. Our idea consists in the extrapolation and interpolation of the epipolar geometry linking the reference views, at the uncalibrated level. With two views we can generate an arbitrary number of synthetic views as the virtual camera moves along a curve. A third view would allow the camera to move on a 2-manifold.

Acknowledgments

Giandomenico Orlandi contributed to this work with inspiring evening discussions.

7. REFERENCES

[1] Cha Zhang and Tsuhan Chen, "A survey on image-based rendering - representation, sampling and compression," Tech. Rep. AMP 03-03, Electrical and Computer Engineering - Carnegie Mellon University, Pittsburgh, PA 15213, June 2003.

[2] O. D. Faugeras and L. Robert, "What can two images tell us about a third one?," in *Proceedings of the European Conference on Computer Vision*, Stockholm, 1994, pp. 485–492.

[3] Leonard McMillan and Gary Bishop, "Plenoptic modeling: An image-based rendering system," in *SIGGRAPH 95 Conference Proceedings*, Aug. 1995, pp. 39–46.

[4] Steven M. Seitz and Charles R. Dyer, "View morphing: Synthesizing 3D metamorphoses using image transforms," in *SIGGRAPH 96 Conference Proceedings*, Aug. 1996, pp. 21–30.

Fig. 1. In each row, the second and the fourth images are the reference ones ($t = 0, t = 1$); the first and the last are extrapolated views ($t = \pm 2$) and the central view is interpolated ($t = 0.5$).

[5] S. Laveau and O. Faugeras, "3-D scene representation as a collection of images and foundamental matrices," Technical Report 2205, INRIA, Institut National de Recherche en Informatique et an Automatique, February 1994.

[6] S. Avidan and A. Shashua, "Novel view synthesis in tensor space," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 1997, pp. 1034–1040.

[7] M. Irani and P. Anandan, "Parallax geometry of pairs of points for 3D scene analysis," in *Proceedings of the European Conference on Computer Vision*, 1996, pp. 17–30.

[8] A. Shashua and N. Navab, "Relative affine structure: Canonical model for 3D from 2D geometry and applications," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 18, no. 9, pp. 873–883, September 1996.

[9] Marc Alexa, "Linear combination of transformations," in *Proceedings of the 29th annual conference on Computer graphics and interactive techniques*. 2002, pp. 380–387, ACM Press.

[10] Andreas Ruf and Radu Horaud, "Projective rotations applied to a pan-tilt stereo head," in *IEEE Conference on Computer Vision and Pattern Recognition*, Fort Collins, Colorado, June 1999, pp. 144–150, IEEE Computer Society Press.

[11] J. Konrad, "Enhancement of viewer comfort in stereoscopic viewing: parallax adjustment," in *SPIE Symposium on Electronic Imaging Stereoscopic Displays and Virtual Reality Systems*, San Jose, CA, January 1999, pp. 179–190.

[12] Q.-T. Luong and T. Viéville, "Canonical representations for the geometries of multiple projective views," *Computer Vision and Image Understanding*, vol. 64, no. 2, pp. 193–229, 1996.

[13] A. Fusiello, S. Caldrer, S. Ceglie, N. Mattern, and V. Murino, "View synthesis from uncalibrated images using parallax," in *12th International Conference on Image Analysis and Processing*, Mantova, Italy, Septermber 2003, IAPR, pp. 146–151, IEEE Computer Society.

[14] T. Viéville, O. Faugeras, and Q.-T. Luong, "Motion of points and lines in the uncalibrated case," *International Journal of Computer Vision*, vol. 17, no. 1, pp. 7–42, Jan. 1996.

[15] J. Shade, S. Gortler, L. He, and R. Szeliski, "Layered depth images," in *SIGGRAPH 98 Conference Proceedings*, 1998.