UNCALIBRATED INTERPOLATION OF RIGID DISPLACEMENTS FOR VIEW SYNTHESIS

A. Colombari, A. Fusiello and V. Murino

Università degli studi di Verona
Dipartimento di Informatica
Strada Le Grazie, 15 - 37134 Verona, Italy
colombari@sci.univr.it, {andrea.fusiello,vittorio.murino}@univr.it

ABSTRACT

In this paper we present a method for novel view synthesis from two uncalibrated reference views. Snapshots of a scene are created as if they were taken from a different “virtual” viewpoint. The relative affine structure is used to describe the geometry of the scene and then to extrapolate and interpolate novel views. The contribution of this paper is an automatic method for specifying the virtual viewpoint needs to be moved. The idea is that the viewer might use a “3D-ness” knob [13] 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.

The rest of the paper is structured as follows. In Section 2, we show the theory necessary to make the paper self-consistent. Section 3 represents the core of the paper. This section describes our approach for specifying virtual viewpoints in an uncalibrated setting. Experimental results concerning synthetic and real scenes are shown and commented in Section 4, and conclusions are drawn in Section 5.

1. INTRODUCTION

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

In the context of novel view synthesis, algorithms based on image interpolation yield satisfactory results [4, 5]. Where no knowledge on the imaging device can be assumed, uncalibrated point transfer techniques utilize image-to-image constraints such as the Fundamental matrix [6], trilinear tensors [7], plane+parallax [8], to re-project pixels from a small number of reference images to a given view. Another way of linking corresponding points is the relative affine structure [9], 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 – with the exception of [10], who assumes that the location of the virtual camera is visible in the reference images – 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 [5]. The framework for interpolation of Euclidean transformations was set forth in [11], whereas the idea of manipulating rigid displacements at the uncalibrated level is outlined in [12], 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 [13] 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.

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 [9].

Given a plane \( \Pi \), with equation \( n^T w = d \), two conjugate points \( m_1 \) and \( m_2 \) are related by

\[
m_2 \sim H_{12} m_1 + e_{21} \gamma_1
\]

where \( H_{12} \) is the collineation induced by the plane \( \Pi \) and \( e_{21} \) is the epipole in the second view. The symbol \( \sim \) means equality up to a scale factor. If the 3D point \( w \notin \Pi \), there is a residual displacement, called parallax. This quantity is proportional to the relative affine structure \( \gamma_1 = \frac{a}{d \kappa_1} \) of \( w \) [9], where \( a \) is the orthogonal distance of the 3-D point \( w \) to the plane \( \Pi \) and \( \kappa_1 \) is the distance of \( w \) from the focal plane of the first camera. Points \( m_2, H_{12} m_1 \) and \( e_{21} \) 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 [9], is the following:

A. given a set of conjugate pairs \((m_1^\ell : m_2^\ell)\) \(\ell = 1, \ldots, m\);
B. recover the epipole \(e_{21}\) and the homography \(H_{12}\) up to a scale factor;
C. choose a point \(m_1^0\) and scale \(H_{12}\) to satisfy
\[m_2^0 \sim H_{12}m_1^0 + e_{21}\]
D. compute the relative affine structure \(\gamma_1^\ell\) from (1):
\[\gamma_1^\ell = \frac{(m_2^\ell \times e_{21})^\top (H_{12}m_2^\ell \times m_2^\ell)}{|m_2^\ell \times e_{21}|^2}.\]
E. specify a new epipole \(e_{31}\) and a new homography \(H_{13}\) (properly scaled);
F. transfer points in the synthetic view with
\[m_3^\ell \sim H_{13}m_1^0 + e_{31}\gamma_1^\ell\] (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 \(e_{31}\) and a new (scaled) homography \(H_{13}\). 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 \(G_{12}\) that links the reference views, \(I_1\) and \(I_2\). The synthetic view \(I_3\) will be constructed in such a way that the pose of the corresponding virtual camera with respect to the reference camera is given by \(G_{12}G_{12} = (G_{12})^2\). This will be then extended to any scalar multiple of \(G_{12}\).

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:
\[\frac{\kappa_2}{\kappa_1} m_2 = H_{12}m_1 + e_{21}\gamma_1\] (4)
and view pair 2-3:
\[\frac{\kappa_3}{\kappa_2} m_3 = H_{23}m_2 + e_{32}\gamma_2.\] (5)

In order to obtain an equation relating view 1 and 3, let us substitute the first into the second, obtaining:
\[\frac{\kappa_3}{\kappa_1} m_3 = H_{23}H_{12}m_1 + (H_{23}e_{21} + e_{32}\frac{d_1}{d_2})\gamma_1\] (6)
By comparing this equation to Eq. (1), we obtain:
\[e_{31} = H_{23}e_{21} + e_{32}\frac{d_1}{d_2}\] (7)
The ratio \(\frac{d_1}{d_2}\) in general is unknown, but if II is the plane at infinity then \(\frac{d_1}{d_2} = 1\) (please note that this is approximately true for planes distant from the camera). Therefore, taking the plane at infinity as II and comparing to Eq. (1) we obtain:
\[H_{\infty 13} = H_{\infty 23}H_{\infty 12} \quad \text{and} \quad e_{31} = H_{\infty 23}e_{21} + e_{32}\] (8)

Hence, we can use \(H_{\infty 13}\) and \(e_{31}\) as defined above, in the transfer equation, Eq. (3). In matrix form Eq. (8) writes:
\[D_{13} = D_{23}D_{12}\] (9)
where
\[D_{ij} \triangleq \begin{bmatrix} H_{\infty ij} & e_{ij} \\ 0 & 1 \end{bmatrix}\] (10)
represents a rigid displacement at the uncalibrated level\(^1\). Consequently, the transfer equation that allows to generate the virtual view \(I_3\), can be re-written:
\[m_3^\ell \sim D_{13} \begin{bmatrix} m_1^\ell \\ \gamma_1^\ell \end{bmatrix}\] (11)

We will now prove that the virtual camera so obtained is displaced from the first one by \(G_{13} = G_{23}G_{12}\). Let
\[G_{ij} \triangleq \begin{bmatrix} R_{ij} & t_{ij} \\ 0 & 1 \end{bmatrix}\] (12)
be a matrix that represents a rigid displacement, where \(R\) is a rotation matrix and \(t\) is a vector representing a translation. We know that composition of rigid displacement correspond to multiplication of such matrices, hence \(G_{13} = G_{23}G_{12}\). In other words, rigid displacements form a group, known as the special Euclidean group of rigid displacements in 3D, denoted by \(SE(3)\). One might conjecture that the uncalibrated rigid displacements form a group as well. Indeed, each element \(D_{ij}\) is the conjugate of an element \(G_{ij} \in SE(3)\) by the matrix \(\tilde{A} = \begin{bmatrix} A & 0 \\ 0 & 1 \end{bmatrix}\):
\[D_{ij} = \begin{bmatrix} A & 0 \\ 0 & 1 \end{bmatrix} G_{ij} \begin{bmatrix} A^{-1} & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} A & 0 \\ 0 & 1 \end{bmatrix}^{-1} \begin{bmatrix} R_{ij} & t_{ij} \\ 0 & 1 \end{bmatrix} A^{-1} = \tilde{A} G_{ij} \tilde{A}^{-1}\] (13)

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 \(H_{\infty 12}\) is fixed, since \(\det(H_{\infty 12}) = 1\) (see [14]). So, point C in the general view synthesis procedure must be replaced with
C. scale \(H_{\infty 12}\) such that \(\det(H_{\infty 12}) = 1\).

\(^1\)Technically, since we assume to know the plane at infinity, this correspond to the affine calibration stratum [14].
As far as point E please note that formulas defined in (8) hold with the
equality sign, hence there are no free scale factors to fix.

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

The same trick cannot be applied to a generic homography
induced by a plane \(P\), essentially because the equation of the plane
is view-dependent. More specifically, if view pair 1-2 and view
pair 2-3 are related by the same rigid displacement, if \(H_{112}\) trans-
fer pair of \(P\) from view 1 to view 2, the same homography will
not transfer correctly points from view 2 to view 3.

Integer exponents provides us with an extrapolation scheme by
discrete steps. However, \(SE(3)\) is also a differentiable manifold,
on which we can make sense of the interpolation between two ele-
ments as drawing the geodesic path between them.
Let us consider, without loss of generality, the problem of inte-
ratitling between the element \(G\) and the identity \(I\). 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 neigh-
borhood of \(I\) into the tangent space to \(SE(3)\) at \(I\). A straight path
in the tangent space emanating from \(0\) is mapped onto a geodesic
in \(SE(3)\) emanating from \(I\) by the exponential map. Hence, the
geodesic path in \(SE(3)\) joining \(I\) and \(G\) is given by

\[
G^t \triangleq \exp(t \log(G)), \quad t \in [0, 1],
\]

More in general, we can define a scalar multiple of rigid transforma-
tions \([11]\):

\[
t \odot G \triangleq G^t = \exp(t \log(G)), \quad t \in \mathbb{R}.
\]

Mimicking the definition that we have done for rigid transforma-
tions, let us define

\[
t \odot D \triangleq D^t = \exp(t \log(D)), \quad t \in \mathbb{R}.
\]

If we use \(D_{14(t)} = t \odot D_{12}\) in the synthesis, as \(t\) varies
we obtain a continuous path that interpolates between the two real
views for \(t < 1\), and extrapolates the seed displacement for \(t > 1\).
In this way we are able to move the uncalibrated virtual camera
continuously on a curve. The parameter \(t\) is the ‘3D-ness’ knob
that we mentioned in the Introduction.

At a calibrated level, this is equivalent to move the camera
along the trajectory \(t \odot G\). Indeed,

\[
D^t = (\hat{A}G\hat{A}^{-1})^t = e^{t \log(\hat{A}G\hat{A}^{-1})} = e^{t \log(G)} \hat{A}^{-1}\]

\[
= A e^{(t \log(G)) \hat{A}^{-1}} = \hat{A} G^t \hat{A}^{-1}
\]

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).

Finally, in order for our method to make sense, we must make
sure that the real logarithm of \(D\) exists. A sufficient condition
for a real invertible matrix \(A\) to have a real logarithm is that \(A\)
has no eigenvalues on the closed negative real axis of the complex
plane \([15]\). \(G\) satisfy the condition because its eigenvalues are
\(\{1, 1, e^{\pm i \theta}\}\) and so does \(D\) because it is similar to \(G\).

4. RESULTS

We performed tests with both synthetic and real images. The for-
er was used to check the extrapolated view produced by the
algorithm against a ground-truth image. The latter to see what is
to be expected from our technique in a real, general situation.

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 implicit-
ly 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 \([16]\).

After aligning the input images with respect to the background
plane, the residual parallax allows to segment off-plane points (for-
ground). From this segmentation we are able to compute the epipoles
and to recover the relative affine structure for a sparse set of fore-
ground points. All these steps are better explained in \([17]\).

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” \([18]\). Pix-
els 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 hom-
ography 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 back-
ground plane is built. Since the foreground could occlude a back-
ground area in both the input images, holes could remain in the
mosaic. These holes are filled by interpolating from the pixel values
on the boundary.

Figure 1 shows results with images generated using OpenGL.
The first two are used as reference images, and the third as ground-
truth. As the reader can notice from the difference image, the error
is limited to few pixels, imputable to approximations introduced in
the computation of the relative affine structures.

In figure 2 some novel snapshots synthesized from a stereo
couple of images taken in “Piazza delle Erbe,” Verona, are shown.
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). Sample movies are available on the Inter-
net\(^3\).

5. CONCLUSION

We presented a technique for the specification of novel viewpoints
in the generation of synthetic views. Our idea consists in the ex-
trapolation 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 cam-
era moves along a curve. A third view would allow the camera to
move on a 2-manifold.

Acknowledgments

Giandomenico Orlandi contributed with inspiring discussions. The
comments of the reviewers helped to improve the paper.

\(^2\)We use the \texttt{roifill} MATLAB function, but any inpainting tech-
nique can be used.

\(^3\)http://www.sci.univr.it/~fusiello/demo/synth.
6. REFERENCES