A single-perspective novel panoramic view from radially distorted non-central images

Rana Molana
Computer and Information Science
University of Pennsylvania
molana@grasp.cis.upenn.edu

Kostas Daniilidis
Computer and Information Science
University of Pennsylvania
kostas@cis.upenn.edu

Abstract

In this paper, we propose an image-based technique for panoramic novel-view generation using three uncalibrated wide-angle images as its input. State of the art in novel view generation presumes the calibration and removal of radial distortion or any other deformation resulting from the geometry of a non-central camera. We propose a method which replaces this calibration with the assumption that the epipole corresponding to the novel viewpoint is at the center of radial distortion and that it is known.

1 Introduction

In novel view synthesis, given an image pair, the intensity of each ray in the novel view is determined by finding corresponding point matches between those images, using the epipolar geometries between the two views and with respect to the novel view. This process requires that the given images obey the standard perspective model, with a single viewpoint and no radial distortion. Such assumptions make the creation of novel panoramas a process of acquiring tens of images, each with relatively narrow field-of-view, with special apparatus, and applying calibration and estimation of circular motion [9].

In this paper, we propose an approach where a novel view can be synthesized from multiple views which might be highly radially distorted, and even non-central, without compensating explicitly for the resulting image deformations. We shed new light on the problem (Fig. 1) of "how a scene looks from a scene point?"[8] by eliminating the need for a reference plane and facilitating synthesis of omnidrectional novel views.

We assume that, for $n \geq 3$ views, the location of the epipole of the novel view in each image is known, and, in the case of radially distorted views, that it is coincident with the centre of radial distortion in the image. This assumption is either enforced (for example by choosing a visible scene point as the centre of projection of the view to be synthesized, and actually fixating on this point in the case of radially distorted views), or else it may naturally be satisfied for certain wide-angle imaging systems (for example multiple views of a spherical mirror). Given such a configuration, the 2D star of virtual rays at the novel viewpoint are mapped to a 1D star of lines in each image, which we model as a projection from $\mathbb{P}^2$ (the domain of the novel view) to $\mathbb{P}^1$ (the angles of the line pencil in image plane). We make use of the well known fact that three projections of a line in space (a virtual line in our case) yield a trifocal constraint, from which we extract the projection matrices required for the view synthesis.

The main contributions of this paper are:

- We can use images with arbitrary radial distortion and even non-central cameras, such as imaging the scene via a spherical mirror.
- We show that novel views can be synthesized without recovering the radial calibration or the mapping from pixels to rays in non-central cameras. All steps for the extraction of projection matrices are either linear or in closed form.

Related work: Related work maybe be divided into two categories: (1) approaches concerning calibration and motion estimation with non-central and/or radially distorted cameras; (2) approaches for novel view synthesis from perspective views. The closest approach to our work is the multiple view geometry of 1D radial cameras by Thirthala and Pollefeys [18, 17] where the 1D trifocal tensor is used for radial distortion calibration. Our euclidean calibration from the 1D trifocal tensor is based on the fundamental papers by Quan and Kanade [10, 12] and Faugeras *et al* [2], while Astrom and Kahl [1] further study ambiguity of motion estimation from 1D projections. Our work differs from [17] in the fact that we directly synthesize a novel view without going through the process of removing radial distortion. For the three view case, whereas a purely rotating camera is assumed in [17], we make the assumption of fixation at the viewpoint of the novel view. By avoiding the distortion calibration step, we do not require the given views to be from central cameras, nor that their radial distortion should be rotationally symmetric, and we can produce a novel view without scene reconstructon. Based on this fact, our technique is only tangentially related to recent proposed methods for calibration of non-central cameras under certain scene constraints [13] and estimation of their relative motion.

Methods to calibrate the radial distortion simultaneously with multiple view geometry, where the rays are assumed to be central and without knowledge of the scene structure, include the division model of Fitzgibbon [4], which assumes known centre of radial distortion (as does [17], and as shall we), while the focus of radial expansion method [6] includes the radial distortion centre in the estimation. Grossman *et al* [5] assume a central camera with pure rotational motion but no rotational symmetry in the distortion.

The approach of Irani *et al* [8] has been our motivation for modelling the novel view as a target-to-source mapping rather than an epipolar transfer [3]. Irani's approach, like ours here, differs from classic view synthesis and morphing [14] where the novel view is close to the recorded views. Several approaches exist in the literature regarding creation of panoramic views from new viewpoints by selection of the appropriate rays from a dense lightfield. The dense lightfield can be constructed as a concentric mosaic [9], as well as with arbitrary motion trajectories [19], after estimating the camera trajectory and registering all rays in the same coordinate system.

2 Problem definition

We adopt the following conventions. The angle θ is defined by $\tan\theta = \frac{v}{u}$, where (u,v) are pixel coordinates. For convenience, we use the following shorthand notation for three homogeneous entities, all of which depend only on θ:

$$\mathbf{l}(\theta) \sim \begin{pmatrix}-\sin\theta & \cos\theta & 0\end{pmatrix}^\top,\ \mathbf{x}(\theta) \sim \begin{pmatrix}\cos\theta & \sin\theta\end{pmatrix}^\top,\ \mathbf{n}(\theta) \sim \begin{pmatrix}-\sin\theta & \cos\theta\end{pmatrix}^\top. \quad (1)$$

By definition, $\mathbf{l}(\theta) \in \mathbb{P}^2$ is a line through the image origin, while $\mathbf{x}(\theta), \mathbf{n}(\theta) \in \mathbb{P}^1$ are in the direction of, and normal to, that same radial line. Evidently, $\mathbf{x}(\theta) \sim (u,v)^\top$.

Novel view synthesis from perspective views: We first consider the simplest case when input images are perspective (pinhole) camera views of the scene and we pick any world point as the optical center of the novel single-perspective view. Assuming correspondences $\mathbf{x}_1$ and $\mathbf{x}_2$ in the two input images and epipolar geometries $\mathrm{F}_{NV,1}$ and $\mathrm{F}_{NV,2}$ with respect to the novel view, a point in the novel view can be synthesized as $\mathbf{x}_{NV} = \mathrm{F}_{NV,1}\mathbf{x}_1 \times \mathrm{F}_{NV,2}\mathbf{x}_2$ [3]. Traditional multiple view geometry techniques [7] may be applied to recover the epipolar geometry. Formulating the reverse projection of this epipolar transfer, each novel view ray can be projected into all original views and a color/intensity assigned to the ray by finding the (single) point correspondence that must occur between those image lines. Irani *et al* [8] use this framework to recover a novel view from an arbitrary scene point using the parallax with respect to a plane with known homography.

Novel view synthesis from radially distorted or axial non-central views: The framework above required pre-calibration and removal of radial distortion from the input views; moreover, removal of non-central distortions would only be possible via knowledge (e.g.reconstruction) of the scene [16]. We relax these assumptions and assume that for every input view the epipole to the novel view is the centre of radial distortion. Hence, any radial distortion will be along the projection $\mathbf{l}$ of a virtual ray $\mathbf{d}$. This can apply to non-central projections when rays of the non-central input view intersect a common axis and that axis intersects the novel viewpoint. Although pointwise there is no perspective projection, a virtual ray $\mathbf{d}$ through such a novel viewpoint projects to a straight line $\mathbf{l}$ through the novel view epipole in the image. We can thus model the projection of the 2D star of virtual rays $\mathbf{d}$ as a 1D line pencil in each input view.

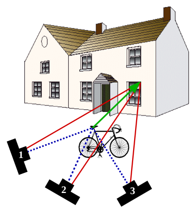

Figure 1. *From three radially distorted, and possibly non-central, views of the scene it is possible to render a novel view of the scene if we assume that the novel view epipole at each of the three input views is the center of radial distortion.*

Although a concurrent novel view epipole and radial distortion centre might seem to require active fixation, some wide angle imaging systems satisfy this automatically. For multiple pinhole camera views of a fixed spherical mirror, captured rays are always non-central [15], yet, regardless of relative camera pose, the sphere centre maps to the radial distortion centre in all views. Thus, choosing the sphere centre as the novel viewpoint fulfils the requirement for epipoles to coincide with the centres of radial distortion for all views. All that remains is to locate the centre of the sphere in each view, achieveable if the perimeter of the mirror is visible in the images. Note that radial distortion will not be rotationally symmetric when the optical axis of the pinhole camera is offset from the centre of the sphere, yet our technique is unaffected since we make no assumptions about the form of the radial distortion.

Assuming a 1D homography between the pencil of planes through each view axis and the pencil of lines in the image, which is indeed the case for most practical "axial" cameras, the mapping of novel view rays ($\mathbb{P}^2$) to image radial line pencil ($\mathbb{P}^1$) is described by a 2×3 homogeneous matrix. (For a perfect pinhole camera this holds at *every* pixel: since it is central, any ray can be picked as the "axis" and a 1D homography is induced on *any* line pencil in the image by the 2D homography of planar perspective projection from pinhole rays to image plane.) We next illustrate by deriving a form for this mapping.

If the rigid body transformation from the novel view coordinate frame to the camera coordinate frame is given by $\mathbf{X}_{\mathrm{C}} = \mathrm{R}\mathbf{X}_{\mathrm{NV}} + \mathbf{T}$, where $\mathrm{R} \in SO(3)$ and $\mathbf{T} \in \mathbb{R}^3$, then the novel view ray with direction $\mathbf{d} \in S^2$ (given in the novel view reference frame) is projected into an input view as the line $\mathbf{l} \in \mathbb{P}^2$ given by

$$\mathbf{l} \sim \mathrm{K}\mathbf{T} \times \mathrm{K}\mathrm{R}\mathbf{d} \quad \text{where K is parameterized as:} \begin{pmatrix} f_u & s & \tilde{u_0} \\ 0 & f_v & \tilde{v_0} \\ 0 & 0 & 1 \end{pmatrix}. \tag{2}$$

The affine transformation K encodes the intrinsic parameters of the mapping from novel view ray ($\mathbf{d}$) to radial line ($\mathbf{l}$). The epipole of the novel view is given by K$\mathbf{T}$; assuming we know its location, we translate our image coordinates to make this epipole our image origin. By doing this, we have set $\mathrm{K}\mathbf{T} \sim (0,0,1)^\top$ in equation (2), and adjusted the variable for the principal point accordingly, giving

$$\mathbf{l} \sim \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} \times \mathrm{K}\mathrm{R}\mathbf{d} \quad \Rightarrow \begin{pmatrix} -\sin\theta \\ \cos\theta \\ 0 \end{pmatrix} \sim \begin{pmatrix} 0 & -1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \mathrm{K}\mathrm{R}\mathbf{d} \Rightarrow \mathbf{x}(\theta) \sim \underbrace{\begin{pmatrix} f_u & s & u_0 \\ 0 & f_v & v_0 \end{pmatrix} \mathrm{R}}_{\mathrm{M}} \ \mathbf{d}. \tag{3}$$

Since M has rank 2, it has a unique RQ-decomposition as the product of a 2×2 homogeneous upper triangular matrix and a 2×3 matrix with orthonormal rows, so the mapping effectively has two intrinsic parameter dofs. Notice that, in a catadioptric case with mirror and pinhole optical axis aligned (i.e., when $\mathbf{T} = (0,0,T_z)^\top$, since the novel view origin also coincides), then $u_0 = v_0 = 0$, and the aspect ratio and skew of the pinhole camera will correspond directly to the two intrinsic parameters of the novel-view-ray-to-radial-line mapping; for that aligned case, any assumptions that can be made about the pinhole camera's aspect ratio and skew parameters can therefore be directly applied to the novel-view-ray-to-radial-line mapping intrinsic parameters.

3 From tri-view correspondences to novel view rays

3.1 Trilinear constraint for radial line correspondences

Given three views, V, V′, V″, we first translate each image, placing the known novel view epipole (assumed to coincide with the radial distortion centre when applicable) at the image origin. Let M, M′, M″ be the 2×3 homogeneous matrices that then map novel view rays to radial lines in respective views. Consider a scene point that is visible in all three views. Let $\mathbf{d} \in S^2$ be the unique novel view ray on which this scene point lies (even if occluded from the novel viewpoint, it still "virtually" coincides with one novel view ray). Let $\mathbf{x}(\theta)$, $\mathbf{x}(\theta')$, $\mathbf{x}(\theta'')$ be the unique radial lines on which the image of this scene point lies in the respective views. Then,

$$\mathbf{x}(\theta) \sim \mathrm{M}\mathbf{d}, \quad \mathbf{x}(\theta') \sim \mathrm{M}'\mathbf{d}, \quad \mathbf{x}(\theta'') \sim \mathrm{M}''\mathbf{d}. \tag{4}$$

Taking the dot product with $\mathbf{n}(\theta), \mathbf{n}(\theta')$ and $\mathbf{n}(\theta'')$ respectively gives

$$\underbrace{\begin{pmatrix} \mathbf{n}(\theta)^\top \mathrm{M} \\ \mathbf{n}(\theta')^\top \mathrm{M}' \\ \mathbf{n}(\theta'')^\top \mathrm{M}'' \end{pmatrix}}_{\mathrm{B}} \mathbf{d} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}. \tag{5}$$

For this equation to hold the determinant of B must equal zero, giving

$$\det(\mathrm{B}) = \mathbf{n}(\theta)^\top \mathrm{M}\,(\mathrm{M}'^\top \mathbf{n}(\theta') \times \mathrm{M}''^\top \mathbf{n}(\theta'')) = 0, \tag{6}$$

which can be written as

$$\begin{pmatrix} -s\theta s\theta' s\theta'' \\ c\theta s\theta' s\theta'' \\ -s\theta c\theta' c\theta'' \\ c\theta c\theta' c\theta'' \\ s\theta s\theta' c\theta'' \\ -c\theta s\theta' c\theta'' \\ s\theta c\theta' s\theta'' \\ -c\theta c\theta' s\theta'' \end{pmatrix}^\top \underbrace{\begin{pmatrix} v_1 \\ v_2 \\ v_3 \\ v_4 \\ v_5 \\ v_6 \\ v_7 \\ v_8 \end{pmatrix}}_{\mathbf{v}} = 0, \quad \text{with } \mathbf{v} \sim \begin{pmatrix} \mathrm{M}(\mathbf{m}'_1 \times \mathbf{m}''_1) \\ \mathrm{M}(\mathbf{m}'_2 \times \mathbf{m}''_2) \\ \mathrm{M}(\mathbf{m}'_1 \times \mathbf{m}''_2) \\ \mathrm{M}(\mathbf{m}'_2 \times \mathbf{m}''_1) \end{pmatrix}, \text{ where } \mathrm{M}' = \begin{pmatrix} \mathbf{m}_1'^\top \\ \mathbf{m}_2'^\top \end{pmatrix} \text{ and } \mathrm{M}'' = \begin{pmatrix} \mathbf{m}_1''^\top \\ \mathbf{m}_2''^\top \end{pmatrix}, \tag{7}$$

where $\sin\theta$ and $\cos\theta$ are denoted as $s\theta$ and $c\theta$ respectively. This $\mathbf{v} \in \mathbb{P}^7$ is the 1D trifocal tensor [11, 18]. If $\mathrm{M}, \mathrm{M}', \mathrm{M}''$ and $\mathbf{d}$ satisfy (5) then for any collineation $\mathrm{H} \in PL(3)$ it will be the case that $\mathrm{MH}, \mathrm{M}'\mathrm{H}, \mathrm{M}''\mathrm{H}$ and $\mathrm{H}^{-1}\mathbf{d}$ also satisfy that constraint as:

$$\mathbf{x}(\theta) \sim \mathrm{M}\mathbf{d} = \mathrm{MH}\,\mathrm{H}^{-1}\mathbf{d}, \quad \mathbf{x}(\theta') \sim \mathrm{M}'\mathbf{d} = \mathrm{M}'\mathrm{H}\,\mathrm{H}^{-1}\mathbf{d}, \quad \mathbf{x}(\theta'') \sim \mathrm{M}''\mathbf{d} = \mathrm{M}''\mathrm{H}\,\mathrm{H}^{-1}\mathbf{d}. \tag{8}$$

This 8 dof homography, H, perfectly accounts for the discrepancy between the 7 dofs of $\mathbf{v}$ and the 15 dofs of the three projection matrices it encodes.

3.2 Estimation of the 1D trifocal tensor

For each scene point that lies on a distinct novel view ray, there will be a different corresponding triplet of $\theta, \theta', \theta''$, and each distinct triplet gives a linear homogeneous constraint on $\mathbf{v}$ of the form in equation (7). Given N distinct corresponding triplets of radial lines across three views, the resulting constraints can be stacked into a matrix equation of the form $\mathrm{D}\mathbf{v} = \mathbf{0}$, where D is a $\mathrm{N}\times 8$ matrix depending only on the measured image angles. If $\mathrm{N} = 7$ then the 1D trifocal tensor, $\mathbf{v}$, can be solved for exactly as the nullspace of D. If $\mathrm{N} > 7$ then $\mathbf{v}$ can be estimated via linear-least squares, for example using SVD. In practice, since this linear estimation requires only 7 correspondence triplets, it is easily and best implemented within a robust framework, such as RANSAC, in order to simultaneously filter the outlier correspondence triplets.

3.3 Determining M, M' and M'' up to a collineation

The method we use to recover the projection matrices up to right-multiplication by an unknown homography was introduced by Quan and Kanade [12], and is briefly outlined here for completeness. Without loss of generality, we parameterize the three projection matrices as:

$$\mathrm{M} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix}, \quad \mathrm{M}' = \left(\begin{pmatrix} c_2 \\ -c_1 \end{pmatrix} \underbrace{\begin{pmatrix} \alpha & \beta \end{pmatrix}}_{\mathbf{a}^\top} \quad \underbrace{\begin{pmatrix} c_1 \\ c_2 \end{pmatrix}}_{\mathbf{c}} \right), \quad \mathrm{M}'' = \begin{pmatrix} \mathbf{m}_1''^\top \\ \mathbf{m}_2''^\top \end{pmatrix}. \tag{9}$$

Since each matrix is homogeneous, scales of M, M', M'' are arbitrary; the 10 parameters in $\mathbf{a}, \mathbf{c}, \mathbf{m}_1'', \mathbf{m}_2''$ thus have 7 dofs which are recovereable from the 7 dof 1D trifocal tensor.

Letting μ denote the homogeneous scale factor, the parameterization of the trifocal tensor in terms of the projection matrices, as given in equation (7), can be written as:

$$\underbrace{\begin{pmatrix} 0 & -c_1 & \beta c_2 & 0 & 0 & 0 & -v_1 \\ c_1 & 0 & -\alpha c_2 & 0 & 0 & 0 & -v_2 \\ 0 & 0 & 0 & 0 & -c_2 & -\beta c_2 & -v_3 \\ 0 & 0 & 0 & c_2 & 0 & \alpha c_1 & -v_4 \\ 0 & 0 & 0 & 0 & -c_1 & \beta c_2 & -v_5 \\ 0 & 0 & 0 & c_1 & 0 & -\alpha c_2 & -v_6 \\ 0 & -c_2 & -\beta c_2 & 0 & 0 & 0 & -v_7 \\ c_2 & 0 & \alpha c_1 & 0 & 0 & 0 & -v_8 \end{pmatrix}}_{\mathrm{G}} \begin{pmatrix} \mathbf{m}''_1 \\ \mathbf{m}''_2 \\ \mu \end{pmatrix} = \mathbf{0}. \quad (10)$$

Since $(\mathbf{m}''_1; \mathbf{m}''_2; \mu) \neq \mathbf{0}$, all the 7×7 minors of G must equal zero, yielding two algebraically independent equations, homogeneous in both $\mathbf{a}$ and $\mathbf{c}$, as follows:

$$\underbrace{\begin{pmatrix} c_1 v_3 - c_2 v_5 & c_1 v_4 - c_2 v_6 \\ c_1 v_7 - c_2 v_1 & c_1 v_8 - c_2 v_2 \end{pmatrix}}_{\mathrm{C}} \underbrace{\begin{pmatrix} c_1^2 + c_2^2 & 0 \\ 0 & c_1 + c_2 \end{pmatrix}}_{\Gamma} \mathbf{a} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}. \quad (11)$$

These can be solved in closed form for $\mathbf{a}$ and $\mathbf{c}$, up to scale in each case. As neither $c_1 = c_2 = 0$ (which makes M' the null matrix) nor $c_1 = -c_2$ (which makes M' rank 1) is possible, it must be that $\det \mathrm{C} = 0$, which gives

$$\det \mathrm{C} = (v_3 v_8 - v_4 v_7) c_1^2 + (v_1 v_4 + v_6 v_7 - v_2 v_3 - v_5 v_8) c_1 c_2 + (v_2 v_5 - v_1 v_6) c_2^2 = 0. \quad (12)$$

In general, this quadratic has two distinct solutions; since the scale of $\mathbf{c}$ directly represents the overall scale of M', we arbitrarily fix it to $|\mathbf{c}| = 1$ for both. Then, substituting each $\mathbf{c}$ solution in equation (11), the respective solutions for $\mathbf{a}$ are found as the nullspace of $\mathrm{C}\Gamma$. The unknown scale of $\mathbf{a}$, corresponds to the arbitrary scale of M, so fix we can fix $|\mathbf{a}| = 1$. In this way, two solutions are found for the matrix M', and for each of these two solutions a unique M'' may be determined from the kernel of G in equation (10).

3.3.1 Discerning the correct solution from the two-fold ambiguity

If we reconstruct the novel view rays for the correspondences, then both solutions will not be within a collineation of the euclidean novel view. Without any further information, it is not possible to distinguish between the two solutions, however in practice many options exist (e.g. a fourth view; information about the scene; an assumption that the radial distortion is constant along conical contours) which would allow for the incompatible solution to be easily identified.

3.4 Estimating the collineation H

When nothing is known regarding the original views' intrinsic parameters or the scene structure, the only option is to pick a collineation that makes the novel view "appear reasonable", for example mapping four of the projective novel rays (reconstructed from correspondences and the projective estimates of M, M', M'') to four desired coordinates.

If information can be gleaned about the scene, or else assumptions made regarding the intrinsic parameters, then sufficient constraints may be derived to determine the particular $\mathrm{H} \in PL(3)$ which gives a euclidean novel view. Details will be case specific; we outline underlying principles for each of these two approaches in the next two subsections.

3.4.1 Using assumptions regarding the original views' intrinsic parameters

Let $\mathrm{M}^{(i)}$, for $i = 1 \ldots n$ views, be the estimated 2×3 projection matrices [1], and let H be the common collineation that upgrades them to the euclidean frame. Consider the following three possible parameterizations for the euclidean $\mathrm{M}^{(i)}\mathrm{H}$:

$$\mathrm{M}^{(i)}\mathrm{H} \sim \bar{\mathrm{K}}^{(i)}\bar{\mathrm{R}}^{(i)} \quad \text{where } \bar{\mathrm{K}}^{(i)} = \begin{pmatrix} \bar{a}_i & \bar{s}_i \\ 0 & 1 \end{pmatrix},\ \bar{\mathrm{R}}^{(i)} \in \mathbb{R}^{2\times3} \text{ with orthonormal rows .} \tag{13}$$

$$\mathrm{M}^{(i)}\mathrm{H} \sim \check{\mathrm{K}}^{(i)}\begin{pmatrix}\check{\mathrm{R}}^{(i)} & \mathbf{t}^{(i)}\end{pmatrix} \quad \text{where } \check{\mathrm{K}}^{(i)} = \begin{pmatrix} \check{a}_i & \check{s}_i \\ 0 & 1 \end{pmatrix},\ \check{\mathrm{R}}^{(i)} \in SO(2),\ \mathbf{t}^{(i)} \in \mathbb{R}^2. \tag{14}$$

$$\mathrm{M}^{(i)}\mathrm{H} \sim \tilde{\mathrm{K}}^{(i)}\mathrm{R}^{(i)} \quad \text{where } \tilde{\mathrm{K}}^{(i)} = \begin{pmatrix} f_{ui} & s_i & u_{0i} \\ 0 & f_{vi} & v_{0i} \end{pmatrix} \text{ and } \mathrm{R}^{(i)} \in SO(3). \tag{15}$$

Unique parameterizations of both forms (13) and (14) will exist for all n euclidean projection matrices [2]. The blatant over-parameterization in (15) is presented as it is more intuitive for the practical case where the final image is from a pinhole camera, perhaps following a mirror reflection, as derived in section 2. Under these parameterizations, the ith camera parameters are related to the common H as follows.

Under parameterization (13) the relation is:

$$\mathrm{M}^{(i)}\mathrm{H}\mathrm{H}^\top\mathrm{M}^{(i)\top} \sim \bar{\mathrm{K}}^{(i)}\underbrace{\bar{\mathrm{R}}^{(i)}\bar{\mathrm{R}}^{(i)\top}}_{I_{2\times2}}\bar{\mathrm{K}}^{(i)\top} = \bar{\mathrm{K}}^{(i)}\bar{\mathrm{K}}^{(i)\top} = \begin{pmatrix} \bar{a}_i^2 + \bar{s}_i^2 & \bar{s}_i \\ \bar{s}_i & 1 \end{pmatrix} \tag{16}$$

This gives two equations per view, relating the elements of $\mathrm{HH}^\top$ to parameters of that view:

$$\frac{\mathbf{m}_1^{(i)\top}\mathrm{HH}^\top\mathbf{m}_1^{(i)}}{\mathbf{m}_2^{(i)\top}\mathrm{HH}^\top\mathbf{m}_2^{(i)}} = \bar{a}_i^2 + \bar{s}_i^2 \qquad \text{and} \qquad \frac{\mathbf{m}_1^{(i)\top}\mathrm{HH}^\top\mathbf{m}_2^{(i)}}{\mathbf{m}_2^{(i)\top}\mathrm{HH}^\top\mathbf{m}_2^{(i)}} = \bar{s}_i \tag{17}$$

Under the parameterization (15) the relation is:

$$\mathrm{M}^{(i)}\mathrm{H}\mathrm{H}^\top\mathrm{M}^{(i)\top} \sim \tilde{\mathrm{K}}^{(i)}\underbrace{\mathrm{R}^{(i)}\mathrm{R}^{(i)\top}}_{I_{3\times3}}\tilde{\mathrm{K}}^{(i)\top} = \begin{pmatrix} f_{ui}^2 + s_i^2 + u_{0i}^2 & f_{vi}s_i + u_{0i}v_{0i} \\ f_{vi}s_i + u_{0i}v_{0i} & f_{vi}^2 + v_{0i}^2 \end{pmatrix} \tag{18}$$

(This is the top 2×2 submatrix of the full $\mathrm{KK}^\top$ - the dual of the image of the absolute conic [7] - for the pinhole camera.) This gives two equations per view, relating the elements of $\mathrm{HH}^\top$ to the parameters of that view, as follows:

$$\frac{\mathbf{m}_1^{(i)\top}\mathrm{HH}^\top\mathbf{m}_1^{(i)}}{\mathbf{m}_2^{(i)\top}\mathrm{HH}^\top\mathbf{m}_2^{(i)}} = \frac{f_{ui}^2 + s_i^2 + u_{0i}^2}{f_{vi}^2 + v_{0i}^2} \qquad \text{and} \qquad \frac{\mathbf{m}_1^{(i)\top}\mathrm{HH}^\top\mathbf{m}_2^{(i)}}{\mathbf{m}_2^{(i)\top}\mathrm{HH}^\top\mathbf{m}_2^{(i)}} = \frac{f_{vi}s_i + u_{0i}v_{0i}}{f_{vi}^2 + v_{0i}^2} \tag{19}$$

Since, H always appears as a symmetric matrix $\mathrm{C} = \mathrm{HH}^\top$, recovery of H amounts to recovering the 5 parameters of C. Once a positive-definite C is recovered, H can be

[1] Thus far, we assumed $n = 3$ views, but $n > 3$ views could, for example, first be processed as triplets comprising one common view, and then registered using a factorization approach [11].

[2] These two parameterizations follow from the RQ decompostions of a 2×3 and 2×2 matrix, respectively, where in both cases the decomposition is unique because the rank is known to be 2. The assumption that the first 2×2 submatrix has rank 2 across all views is without loss of generality, since the euclidean projection matrices are only up to a similarity, and that similarity can always be chosen to ensure this fact.

found via Cholesky decompositon. In particular, if we know that the parameter ratios are the same across several views, then with just 4 such views we can recover both H and the two common ratios from the RHS of (19). For the case with several randomly posed pinhole camera views of a fixed spherical mirror, (u_{0i}, v_{0i}) will vary for each view, precisely because we translate each image differently to make each epipole the origin. However, since the offset is known for each view, we can write all the (u_{0i}, v_{0i}) in terms of a common unknown principal point and again, provided we have sufficient views, and provided the pinhole camera intrinsics are constant, we can estimate the collineation, albeit non-linearly. If the principal point can be assumed concurrent with the centre of radial distortion (i.e., camera and mirror axes aligned), and pinhole camera skew assumed zero, then 3 such views are sufficient and the solution for C is linear.

3.4.2 Using assumptions regarding the scene

Given projective estimates M, M′, M″, we can projectively reconstruct novel view rays for the correspondence points. Finding the collineation that maps these projective novel rays to a euclidean frame is equivalent to the problem of calibration from a single view, which can be approached variously [7]. If the scene allows identification of two vanishing points (thus horizon of a plane) in the projectively reconstructed novel view, we can upgrade to an unknown affinity. Additionally, if the circular points can be identified, the calibration may be updated up to a similarity.

3.5 Overall algorithm: recovering novel view epipolar geometry

Given three views of a scene such that the novel view epipole in each image is known (and coincides with the centre of radial distortion, where applicable):

1. For each view, translate pixel coodinates so the novel view epipole is the image origin.

2. Compute $N \geq 7$ distinct line correspondence triplets $\mathbf{x}(\theta), \mathbf{x}(\theta'), \mathbf{x}(\theta'')$ from point correspondences across the three images.

3. Estimate the best-fit $\mathbf{v}$, via SVD, for the ≥ 7 homogeneous equations of the form in (7), after first filtering outlier correspondences using a robust estimator such as RANSAC.

4. Substitute $\mathbf{v}$ in equation (12) and solve the quadratic to obtain two homogeneous solutions for $\mathbf{c}$. Set $|\mathbf{c}| = 1$ for both solutions.

5. Substitute each of the two solutions for $\mathbf{c}$ in equation (11) and solve to obtain two corresponding homogeneous solutions for $\mathbf{a}$. Set $|\mathbf{a}| = 1$ for both solutions.

6. For each of the two $(\mathbf{a}, \mathbf{c})$-solutions, substitute $\mathbf{v}, \mathbf{a}, \mathbf{c}$ to build the 8×7 matrix G in (10). Extract $(\mathbf{m}_1'', \mathbf{m}_2'')$ from the nullspace of G.

7. Build the two solutions for the projection matrices M, M′, M″ from the two solutions for $(\mathbf{a}, \mathbf{c}, \mathbf{m}_1'', \mathbf{m}_2'')$, according to the parameterization in (9).

Steps for discerning the correct solution from the two-fold ambiguity, and for recovering the collineation to a euclidean novel view, are case specific, as detailed in sections 3.3.1 and 3.4.

4 Rendering the novel view

Let $\mathrm{M}^{(i)}$, for $i = 1 \ldots n$ views, be the 2×3 projection matrices estimated from methods in section 3. If we had enough information to determine the collineation, these projection matrices refer to the euclidean frame and a euclidean novel view will ensue; if not, techniques in this section will lead to a projective novel view.

A sparse novel view from correspondences Let $(\theta_1, \theta_2, \ldots, \theta_m)$ be a correspondence tuplet across $m \geq 2$ of the n views. Then,

$$\begin{pmatrix} \mathbf{n}(\theta_1)^\top \mathrm{M}^{(1)} \\ \vdots \\ \mathbf{n}(\theta_m)^\top \mathrm{M}^{(m)} \end{pmatrix} \mathbf{d} = \mathbf{0}. \tag{20}$$

If $m = 2$ the corresponding novel view ray, $\mathbf{d}$, is the cross-product. For $m > 2$ the nullspace (via SVD) gives $\mathbf{d}$. Using inlier correspondences (including two-view correspondences) from the initial 2D matching for trifocal tensor estimation results in a forward version of novel view synthesis, entailing a short run time but rendering a sparse novel view.

A dense novel view via guided matching Establishing a dense novel view means iterating over all rays to be synthesized, projecting each into all input views. The ray color/intensity can be set from the most photoconsistent pixel along the corresponding radial lines. Ordering should be considered when matching, as multiple points on the same novel ray could be visible in the original views (with a fixed spherical mirror case this not likely: parallax between views due to a non-central caustic [15] would be very small).

5 Preliminary simulations and experiments

Simulations for computing novel view epipolar geometry, using real 3D scene points projected into four synthetic views, are currently being undertaken. A preliminary remark is that correspondences should span > 90 degrees in at least one image for stability. Thus, although the framework itself does not preclude configurations where the novel view epipole/centre of radial distortion is on the perimeter, or even outside the frame, of *all* images (e.g., oblique pinhole views of a spherical mirror slice), such configurations are not ideal. In practical imaging, the radial distortion centre does usually lie within the image, and since wide-angle imaging captures many features, tri-view point correspondences are likely to span the full range of θ in at least one view, making estimation stable. Implementation with a sequence of shots of a spherical mirror, such as in Fig. 2, is future work, but note that correspondences have wide θ span in all images, and that even just three views of a spherical mirror provide more than enough information for a full panoramic novel view, as each visible scene point is captured by at least two images.

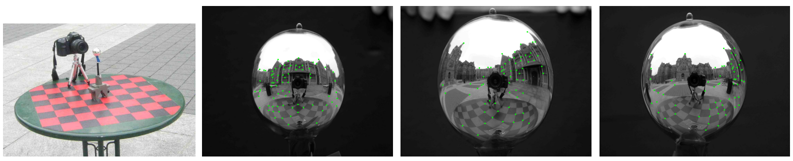

Figure 2. *The recording setup on the left and the three input views.*

6 Conclusion

This paper has proposed a framework for rendering a single-perspective novel panoramic view from radially distorted non-central images, when the epipole of the novel view can be assumed to coincide with the centres of radial distortion in all views. This condition is *automatically* met by multiple views of a fixed spherical mirror using a pinhole camera (whatever the relative pose), and that is a primary practical use for this framework.

For other wide-angle imaging techniques (using a hand-held camera with a fish-eye lens, say!), precise fixation by the centres of radial distortion on a single scene point in space will be prone to some error, and analysis of sensitivity of the epipolar geometry estimation to errors in the fixation and/or errors in the assumed centres of radial distortion is currently being undertaken, in order to quantify framework limitations. It should be noted that state-of-the-art techniques for generating novel panoramic views typically assume circular motion, or pure rotation, and so our requirement that the views should fixate on a novel viewpoint is actually *less* restrictive.

Rendering a dense novel view will depend on the success of guided matching along the corresponding radial lines. Stereo rectification or, more appropriately, rectification via a homography to one of the original views as in [8], is not a possibility because the intra-view epipolar geometry is not recovered (since no model is assumed for the radial distortion). Stereo rectification to the novel view will not be useful for typical configurations, where the novel view epipole lies within the images. Exploration of techniques for this 1D guided matching is therefore of interest for future work.

References

[1] Kalle Astrom and Fredrik Kahl. Ambiguous configurations for the 1d structure and motion problem. *Journal of Mathematical Imaging and Vision*, 18(2):191–203, 2003.

[2] O. Faugeras, L. Quan, and P. Sturm. Self-calibration of a 1d projective camera and its application to the self-calibration of a 2d projective camera. *PAMI*, 22(10):1179–1185, 2000.

[3] O. Faugeras and L. Robert. What can two images tell us about a third one? In *ECCV*, pages 485–492, 1994.

[4] A. W. Fitzgibbon. Simultaneous linear estimation of multiple view geometry and lens distortion. In *CVPR*, pages 125–132, 2001.

[5] E. Grossmann, E. Lee, P. Hislop, D. Nister, and H. Stewenius. Are two rotational flows sufficient to calibrate a smooth non-parametric sensor? In *CVPR*, pages 1222–1229, 2006.

[6] R. Hartley and S.B. Kang. Parameter-free radial distortion correction with centre of distortion estimation. In *ICCV*, pages 1834–1841, 2005.

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

[8] M. Irani, T. Hassner, and P. Anandan. What does the scene look like from a scene point? In *ECCV*, pages 883–897, 2002.

[9] G. Jiang, Y. Wei, L. Quan, H. Tsui, and H.Y. Shum. Outward-looking circular motion analysis of large image sequences. *PAMI*, 27(2):271–277, 2005.

[10] L. Quan. Self-calibration of an affine camera from multiple views. *IJCV*, 19(1):93–105, 1996.

[11] Long Quan and Takeo Kanade. A factorization method for affine structure from line correspondences. In *CVPR*, page 803, 1996.

[12] Long Quan and Takeo Kanade. Affine structure from line correspondences with uncalibrated affine cameras. *PAMI*, 19(8):834–845, 1997.

[13] Srikumar Ramalingam, Peter F. Sturm, and Suresh K. Lodha. Theory and calibration for axial cameras. In *ACCV*, pages 704–713, 2006.

[14] S. M. Seitz and C. R. Dyer. View morphing. In *SIGGRAPH*, pages 21–30, 1996.

[15] R. Swaminathan, M. D. Grossberg, and S. K. Nayar. Non-Single Viewpoint Catadioptric Cameras: Geometry and Analysis. *IJCV*, 66(3):211–229, Mar 2006.

[16] R. Swaminathan, M.D. Grossberg, and S.K. Nayar. A Perspective on Distortions. In *CVPR*, volume II, pages 594–601, Jun 2003.

[17] S. Thirthala and M. Pollefeys. Multi-view geometry of 1d radial cameras and its application to omnidirectional camera calibration. In *ICCV*, pages 1539–1546, 2005.

[18] S. Thirthala and M. Pollefeys. The radial trifocal tensor: A tool for calibrating the radial distortion of wide-angle cameras. In *CVPR*, pages 321–328, 2005.

[19] A. Zomet, D. Feldman, S. Peleg, and D. Weinshall. Mosaicing new views: the crossed-slits projection. *PAMI*, 25(6):741–754, June 2003.