Lens Distortion Recovery for Accurate Sequential Structure and Motion Recovery

Kurt Cornelis*, Marc Pollefeys*, and Luc Van Gool

K. U. Leuven, ESAT-PSI
Kasteelpark Arenberg 10
B-3001 Leuven-Heverlee, Belgium
firstname.lastname@esat.kuleuven.ac.be
http://www.esat.kuleuven.ac.be/~kcorneli
http://www.esat.kuleuven.ac.be/~pollefey

Abstract. Lens distortions in off-the-shelf or wide-angle cameras block the road to high accuracy Structure and Motion Recovery (SMR) from video sequences. Neglecting lens distortions introduces a systematic error buildup which causes recovered structure and motion to bend and inhibits turntable or other loop sequences to close perfectly. Locking back onto previously reconstructed structure can become impossible due to the large drift caused by the error buildup. Bundle adjustments are widely used to perform an ultimate post-minimization of the total reprojection error. However, the initial recovered structure and motion needs to be close to optimal to avoid local minima. We found that bundle adjustments cannot remedy the error buildup caused by ignoring lens distortions. The classical approach to distortion removal involves a preliminary distortion estimation using a calibration pattern, known geometric properties of perspective projections or only 2D feature correspondences. Often the distortion is assumed constant during camera usage and removed from the images before applying SMR algorithms. However, lens distortions can change by zooming, focusing and temperature variations. Moreover, when only the video sequence is available preliminary calibration is often not an option. This paper addresses all fore-mentioned problems by sequentially recovering lens distortions together with structure and motion from video sequences without tedious pre-calibrations and allowing lens distortions to change over time. The devised algorithms are fairly simple as they only use linear least squares techniques. The unprocessed video sequence forms the only input and no severe restrictions are placed on viewed scene geometry. Therefore, the accurate recovery of structure and motion is fully automated and widely applicable. The experiments demonstrate the necessity of modeling lens distortions to achieve high accuracy in recovered structure and motion.

Key words: Structure from motion, calibration, lens distortion recovery, high accuracy, sequential.

* Kurt Cornelis and Marc Pollefeys are respectively research assistant and postdoctoral fellow of the Fund for Scientific Research - Flanders(Belgium)(F.W.O. - Vlaanderen)

1 Introduction

1.1 Previous Work

Much research exists that acknowledges the importance of modeling lens distortions. Most papers determine lens distortions using calibration patterns [2, 3, 7, 13–15], known geometric properties of perspective projections [2, 4, 8, 9, 11, 12] or 2D feature correspondences [5, 9, 10, 17]. After this pre-calibration, the distortions are often assumed constant during the remainder of the camera usage, a valid assumption if no zooming or focusing is performed. However, when only the video sequence is available a preliminary calibration is often impossible and other ways to recover lens distortions are needed.

Lens distortions are mostly considered after the application of an ideal pinhole projection model. Some work also exists to include distortions implicitly in projection equations using intermediate parameters without physical meaning [14]. The extraction of the distortion parameters from the latter, e.g. to undo or add the same distortion to computer generated graphics which need to be incorporated in Augmented Reality, is often difficult and not well conditioned. This is due to the strong coupling between intrinsic, extrinsic and distortion parameters. As several authors [2, 9, 10, 15] stated, this coupling can result in unacceptable variance of the recovered parameters. Therefore, a decoupling of the projection equations in a part modeling ideal pinhole projection and a part modeling lens distortions is used by [4, 15].

This paper is most closely related to [15] but presents a new way to sequentially determine lens distortions together with structure and motion from a video sequence without preliminary calibration using calibration patterns or specific geometric scene properties. Due to this sequential nature, the lens distortions are also allowed to change over time as can happen in reality.

1.2 Overview

A first section will describe the camera projection model. Subsequently, the estimation of all parameters given 3D-2D correspondences is explained. The model consists of a pinhole projection part and a lens distortion part. The estimation of both are realized iteratively, each described in a separate section. This estimation needs initialization which is considered in a following section. Experiments demonstrate the importance of modeling distortions, we finish with a short summary and propose future research topics.

2 The Camera Model

The model describing the projection process from 3D scene points to 2D image coordinates consists of several sequential steps, shown in Figure 1. First, a projection takes place according to the classical pinhole model.

$$m_p \sim \mathbf{P}_p M_r \tag{1}$$

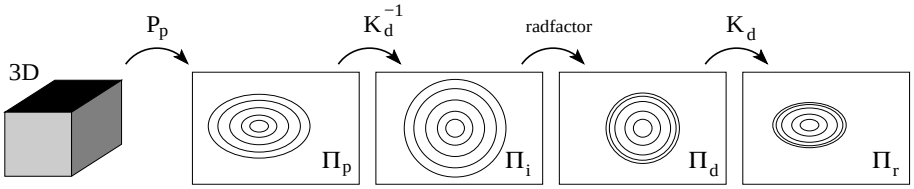

Fig. 1. The camera model from left to right: First, the 3D space is projected onto an image plane Π_p with a pinhole camera. Next, this plane is transformed with $\mathbf{K}_d^{-1}$ to yield the ideal image plane Π_i. In the ideal image plane the radial distortion factor is applied to give coordinates in the distorted image plane Π_d. Finally, this plane is transformed back using $\mathbf{K}_d$ to end up in the real image plane Π_r.

$$\mathbf{P}_p = \mathbf{K}_p[\mathbf{R}^T| - \mathbf{R}^T t]$$
$$\mathbf{K}_p = \begin{bmatrix} f_p & s_p & u_p \\ 0 & r_p f_p & v_p \\ 0 & 0 & 1 \end{bmatrix}$$

in which $M_r = (X_r, Y_r, Z_r, W_r)^T$ and $m_p = (x_p, y_p, w_p)^T$ are the homogeneous coordinates of the 3D point and its projection onto the pinhole image plane Π_p. $\mathbf{P}_p$ is a 3×4 matrix and '$\sim$' denotes that (1) is valid up to a scale factor. $\mathbf{P}_p$ can be decomposed in internal and external calibration parameters. $\mathbf{K}_p$ is called the calibration matrix which contains the focal length f_p, the pixel aspect ratio r_p, the skew s_p and the principal point (u_p, v_p). $\mathbf{R}$ and t determine the rotation and translation in world coordinates.

Next, the distortions transforming the ideal pinhole image into an image that conforms more to real images are modeled. Various distortions exist [16] but radial distortion, described by the following model, is the most prominent.

1. Using a distortion calibration matrix similar to $\mathbf{K}_p$ transform the pinhole image coordinates m_p to the ideal image plane Π_i where the radial distortion center equals $(0, 0)$.

$$m_i = \mathbf{K}_d^{-1} m_p \tag{2}$$
$$\mathbf{K}_d = \begin{bmatrix} f_d & s_d & u_d \\ 0 & r_d f_d & v_d \\ 0 & 0 & 1 \end{bmatrix}$$

$\mathbf{K}_d$ can differ from $\mathbf{K}_p$ as cameras can be recovered in a projective framework in which $\mathbf{K}_p$ has no physical meaning. While radial distortions take place around the physical optical axis, the principal point (u_p, v_p) cannot be used as the distortion center. $\mathbf{K}_d$ is therefore considered the Euclidean version of $\mathbf{K}_p$. As in modern cameras the skew is negligible and aspect ratio is practically 1, we can fix s_d and r_d to 0 and 1 respectively while keeping the distortion center (u_d, v_d) and focal length f_d variable. As shown in [4,

9], the latter allows to model de-centering distortions together with radial distortions. Note, however, that all parameters could be made variable if desired.

2. Next, a radial distortion is applied to m_i to yield the distorted image coordinates m_d in the distorted image plane Π_d.

$$\begin{aligned} m_d &= m_i \times radfactor \\ radfactor &= \sum_{j=0}^{n} k_j r_i^{2j} \\ \text{with } r_i &= \sqrt{x_i^2 + y_i^2} \end{aligned} \tag{3}$$

where k_j are the radial parameters. The model for $radfactor$ is common in literature except that we take k_0 not to equal 1 but make it act as a parameter for a reason later explained. A decreasing $radfactor$ for increasing r_i introduces barrel distortion; otherwise it introduces pincushion distortion.

3. Finally, the distorted coordinates m_d are transformed with $\mathbf{K}_d$ to yield the final real image coordinates m_r in the real image plane Π_r.

$$m_r = \mathbf{K}_d m_d \tag{4}$$

An ambiguity exists between the radial parameters k_j and $\mathbf{K}_d$. Scaling f_d and s_d with sc results in scaling the ideal image plane Π_i around the distortion center with sc^{-1}. Therefore, $radfactor$ has to remain unchanged to obtain identical m_r after distortion. In equation (3) only r_i is scaled, but an appropriate change of k_j can compensate this, hence the ambiguity. To resolve the ambiguity f_d and s_d will be scaled such that every m_i will lie in the range $[-1, 1]$ in Π_i. This will condition algorithms for numerical stability.

Above we stated that strong parameter coupling can result in unacceptable variance of the recovered parameters. This has to be clarified as an ambiguity is a perfect coupling. Suppose two parameters describe an ambiguity, as in $a \times b = c$ where c is a constant an increase in a can be compensated by a decrease in b. The camera model estimation, explained in the following sections, consists of separate steps in which some parameters are held constant at each step. Therefore, fixing one parameter during a step automatically determines the value of the other parameter with which it forms the ambiguity, avoiding any danger of large parameter variances.

3 Camera Model Estimation

During sequential Structure and Motion Recovery from video sequences we advance each time by calculating the camera pose for the current frame given 3D-2D correspondences (M_r, m_r), as explained in [1]. Given the 3D-2D correspondences, we wish to minimize the following error:

$$\min_{\mathbf{P}_p, \mathbf{K}_d, k_j} \sum_l (m_r^l - proj(M_r^l))^2$$

in which $proj()$ denotes the total projection, described in section 2, and $\mathbf{P}_p, \mathbf{K}_d$ and k_j are the parameters to be optimized. The error is the residual reprojection error in the real image which forms the natural goal for minimization as this is the only error visible to human observers.

The camera model consists of two parts; a first part models a pinhole projection; a second part describes lens distortions. We therefore opted to perform the error minimization as an iterative process where each part is minimized while the other is kept constant. As stated above, the coupling between all parameters is strong if solved for in a single global optimization. The decoupling in this multi-step iterative procedure reduces this problem and allows to use simple linear least squares techniques. The iterative procedure has the following lay-out:

1. Initialize the distortion parameters $\mathbf{K}_d$ and k_j.
2. Given the distortion parameters and m_r, one can compute m_p. In Figure 1 this corresponds to a motion from right to left towards the plane Π_p.
3. Given (M_r, m_p) correspondences estimate $\mathbf{P}_p$ minimizing the residual reprojection error.
4. Given the projection matrix $\mathbf{P}_p$ and M_r, one can compute m_p. In Figure 1 this corresponds to a motion from left to right towards the plane Π_p.
5. Given (m_p, m_r) correspondences determine $\mathbf{K}_d$ and k_j minimizing the residual reprojection error.
6. Return to step 2 until convergence.

The initialization step 1 will be explained in section 4. The following sections clarify step 5 and step 3 respectively.

3.1 Lens Distortion Estimation

At this point a previous best pinhole camera matrix $\mathbf{P}_p$ has been determined. Given (m_p, m_r) correspondences, we look for the distortion parameters $\mathbf{K}_d$ and k_j minimizing the following residual reprojection error:

$$\min_{\mathbf{K}_d, k_j} \sum_l (m_r^l - distort(m_p^l))^2 \tag{5}$$

in which $distort()$ determines the second part (lens distortions) of the camera model. The residual is expressed in the real image plane Π_r, the only plane in which we can finally see the errors as all other image planes Π_p, Π_i and Π_d are virtual. Substituting equations (2) and (3) in (4) we get nonlinear equations in the distortion parameters. However, given constant m_i, we note that fixing $\mathbf{K}_d$ gives linear equations in k_j. Vice versa, taking k_j constant yields linear equations in the elements of $\mathbf{K}_d$. Therefore, another iterative solution surfaces:

1. Use the current $\mathbf{K}_d$ and $\mathbf{P}_p$ to form a compound camera matrix which projects a 3D point M_r directly onto the ideal image plane Π_i: $\mathbf{P}_i = \mathbf{K}_d^{-1}\mathbf{P}_p$. Use $\mathbf{P}_i$ to project all M_r to their corresponding m_i which from now on are assumed constant. Note that assuming m_i constant is equal to fixing $\mathbf{P}_i$. In the following steps $\mathbf{K}_d$ will change, requiring $\mathbf{P}_p$ to compensate for this change to keep $\mathbf{P}_i$ constant. The altered $\mathbf{P}_p$ will be determined in step 6.

2. Given $\mathbf{K}_d$ and equations (3) and (4) we can determine k_j minimizing (5) with linear least squares techniques.
3. Using k_j and equation (3) we can distort all m_i to their corresponding m_d in the distorted image plane Π_d.
4. Given m_d and equation (4) we can determine $\mathbf{K}_d$ minimizing (5) with linear least squares techniques.
5. Return to step 2 until convergence.
6. Because m_i and therefore $\mathbf{P}_i$ were assumed constant and $\mathbf{K}_d$ changed during iteration, we update $\mathbf{P}_p$ by extracting the new $\mathbf{K}_d$ from $\mathbf{P}_i$: $\mathbf{P}_p = \mathbf{K}_d\mathbf{P}_i$

The use of parameter k_0, in conventional radial distortion modeling taken to be 1, will now be explained. The camera matrix $\mathbf{P}_p$ used in step 1 will have been estimated before the lens distortions are updated in step $2-5$. Therefore, $\mathbf{P}_p$ will have been estimated based on an image where radial distortion can be under- or overestimated. This leads to a $\mathbf{P}_p$ predicting well the 3D-2D correspondences (M_r, m_r) on a certain circle around the distortion center in Π_r. But for barrel lens distortions it will overestimate the predicted positions of m_r outside this circle and underestimate them inside it and vice versa for pincushion distortions, as shown in Figure 2. These over- and underestimations will be compensated for by the estimation of the distortion parameters in steps $2-5$. However, note that the estimated $radfactor$ should almost equal 1 on the circle for which $\mathbf{P}_p$ already fits best. By fixing k_0 to 1 we also demand $radfactor$ to equal 1 at the distortion center. As Figure 2 shows, this constrains $radfactor$ (3) to be a non-monotone function. However, we know that radial distortions in real lenses have a more or less monotone function. Therefore, it is better to allow k_0 to be a parameter, representing a scaling of the pinhole image plane Π_p and the camera matrix $\mathbf{P}_p$ around the distortion center to achieve a monotone $radfactor$ function as shown in Figure 3.

3.2 Pinhole Camera Estimation

In this section the distortion parameters $\mathbf{K}_d$ and k_j are kept constant and the estimation of camera matrix $\mathbf{P}_p$ using 3D-2D correspondences (M_r, m_p) is considered. Given m_r and the estimated distortion parameters we can undo the distortion to calculate their corresponding m_p. The camera matrix $\mathbf{P}_p$ which minimizes

$$\min_{\mathbf{P}_p} \sum_l (m_p^l - pinhole(M_r^l))^2 \tag{6}$$

can be found by iteratively re-weighted least squares minimization [6] given the previously estimated camera matrix $\mathbf{P}_p$ as initialization. $pinhole()$ represents the pinhole projection model, described in (1), and the residual error (6) is an error living in the pinhole image plane Π_p. Since the only image plane visible to human observers is the real image plane Π_r, the residuals should be transferred to the latter as also noted by [12]. Equation (3) shows that in the neighborhood of m_p the ideal image plane and the pinhole image plane is scaled with

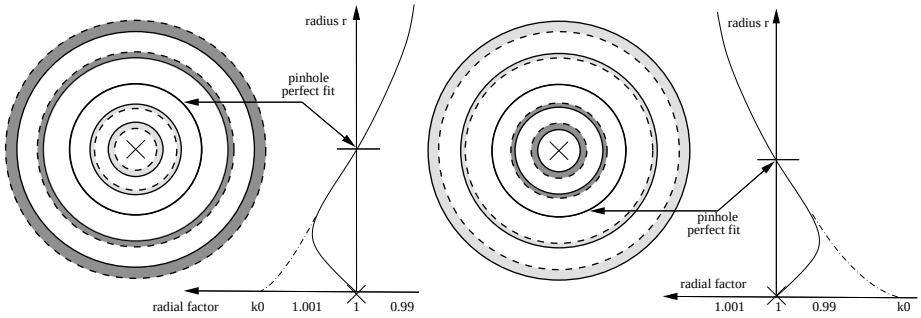

Fig. 2. Left: If the barrel distortion is underestimated the pinhole camera model fits perfectly on a certain circle (thick) around the distortion center but underestimates the distortion outside this circle (dark gray) and overestimates it inside the circle (light gray). The dotted circles are predicted by the pinhole model, the solid circles are the real distorted ones. The estimated radial factor will try to compensate these over- and underestimations. Right: Same but now for pincushion distortion. In both examples the radial factor is forced to create a bump in its curvature (solid curve) while the best match would be the dash-dotted radial factor curve.

$radfactor \sum_{j=0}^{n} k_j r_i^{2j}$ to get to the real image plane. Therefore the residual of m_p is scaled with exactly this factor to calculate its counterpart in the real image plane. The new error measure then becomes:

$$\min_{\mathbf{P}_p} \sum_l [radfactor(m_p^l) * (m_p^l - pinhole(M_r^l))]^2 \tag{7}$$

4 Initialization

The iterative multi-step estimation of all projection parameters needs an initial starting point. Using a sequential Structure and Motion Recovery methodology the projection parameters of each frame are estimated while running through the video. Considering the time-continuity, we can take the previous frame's distortion parameters as a starting point for the current frame. To initialize the sequential recovery of structure, motion and distortion we use the same method as described in [1] where two initial camera matrices $\mathbf{P}_p$ are determined by decomposing a Fundamental Matrix. This decomposition assumed the Fundamental Matrix to be estimated between two images with zero lens distortion and therefore the initial distortion parameters for these cameras can be taken as:

$$\mathbf{K}_d = \begin{bmatrix} f_d & 0 & u_d \\ 0 & f_d & v_d \\ 0 & 0 & 1 \end{bmatrix} \tag{8}$$

$$k_0 = 1 \text{ and } \forall j \neq 0 : \; k_j = 0 \tag{9}$$

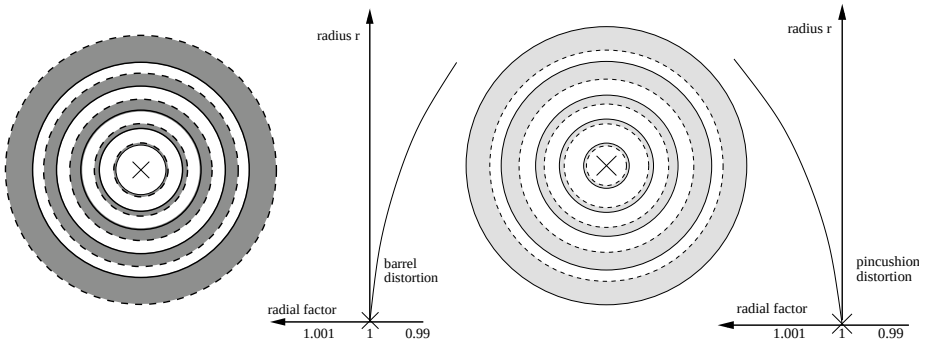

Fig. 3. Left: Same situation as in Figure 2 but now the circles predicted by the pinhole model are up-scaled in the image plane Π_p so that the distortion is an overestimation everywhere. This makes it possible for the radial factor to be a monotone function of the radius which is more natural for real lenses. Right: The same but now the dotted circles predicted by the pinhole camera model are down-scaled.

The initial aspect ratio r_d and skew s_d are respectively taken 1 and 0 which is reasonable for real cameras. The initial distortion center (u_d, v_d) equals the image center and the initial focal length f_d is chosen such that the image corners form a bounding box in the ideal image plane Π_i which lies in the $[-1, 1]$ range for conditioning numerical algorithms.

Given a real camera with lens distortions, we therefore start from an initialization which is not correct as both initial cameras are distortion free. However, minimizing the reprojection error for each frame using the supplied distortion parameters to remove any systematic error, the distortion parameters will most often converge to the real ones. At this point we cannot provide any mathematical proof but the statement is backed up with extensive simulations on artificial and real video sequences which show the relevance of modeling lens distortions. When we dispose of the video camera we could use any available off-line technique that uses calibration patterns to find better initial values for the distortion parameters.

5 Experimental Results

We conducted experiments on artificial and real-life video sequences to test the usefulness of sequential modeling lens distortions during Structure and Motion Recovery. First, an artificial sequence (500 frames, image resolution 720×576) without lens distortions was created. The scene consisted of boxes positioned at different depths. Figure 4 shows the recovered Structure and Motion when no radial distortion was estimated. It corresponds very well with the ground truth. Next, we artificially added barrel lens distortions, which moved points in the image corners with 25 pixels from their original position. At first we did not

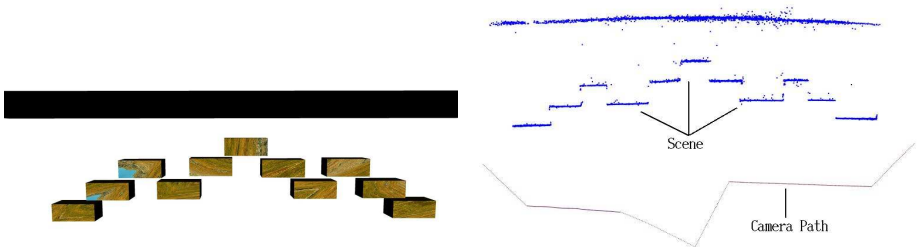

Fig. 4. Left: Top view of the scene structure. Several boxes are placed at different depths. Right: The recovered structure (upper dots) and motion (lower line) from an artificial video sequence without radial distortions.

Fig. 5. Left: The recovered structure and scene when radial distortion was present but not modeled by the program. Right: The recovered structure and motion when the radial distortion was also modeled as described in this paper. The camera moved from right to left. As the system only converged after 200 frames the part of the structure on the far right still shows some residual bending.

try to recover the distortion which introduced a systematic error buildup in the recovered Structure and Motion. Figure 5 shows how the unmodeled radial distortion bends the recovered structure and motion. When we also tried to recover the lens distortions the distortion parameters converged after approximately 200 frames (Figure 6) and the recovered structure and motion resembles more the ground truth as also shown in Figure 5.

Next, a second artificial sequence (image resolution 720×576) was made to investigate the error buildup with or without modeling lens distortions. It consisted of a turntable sequence of a teapot. Figure 7 shows the first frame of the original and radial distorted versions. A single round trip counts 100 frames. We performed ten round trips and therefore the total sequence consisted of 1000 frames. Due to periodicity the frames whose frame number are equal modulo 100 are the same. The error buildup during sequential Structure and Motion Recovery is measured as the projection errors in cameras that are supposed to be the same due to this periodicity. Equal cameras are supposed to project 3D

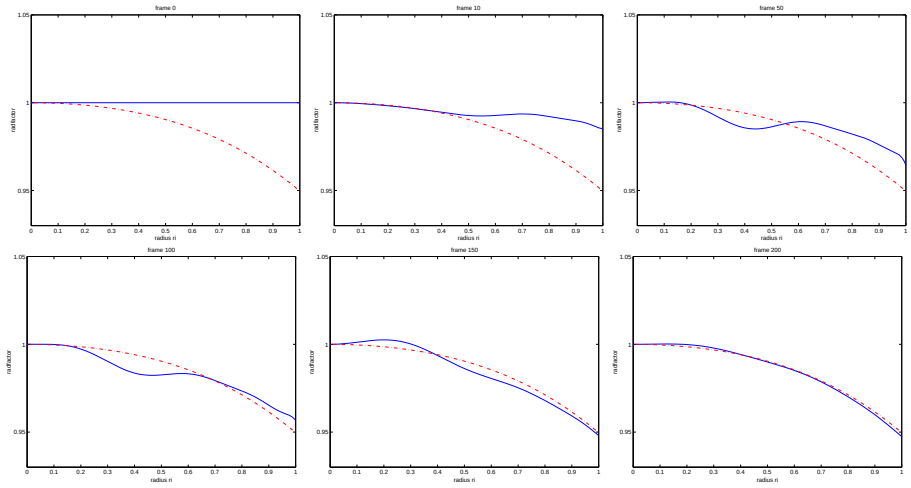

Fig. 6. From left to right [frames 0,10,50,100,150 and 200]: The convergence of the estimated radial factor (solid curve) towards the ground truth (dash-dot). The abscissa represents the radius in the ideal image plane Π_i. The ordinate represents the radial distortion factor *radfactor*.

scene points onto the same image locations. However, this is not the case when error buildup is present which can therefore be measured as the average projection error between 'equal' cameras. This projection error, expressed in pixels, is calculated between cameras separated by one round trip (relative error buildup). Artificial sequences were made having no distortion, barrel distortion and pincushion distortion. For each different kind of distortion the influence of the number of estimated radial distortion parameters was investigated (the number of parameters are counted excluding parameter k_0 to correspond to more common conventions). Table 1 displays the results from which several conclusions can be drawn. First, when radial distortion (barrel or pincushion) is present, not modeling it leads to a large error buildup. When modeling the radial distortion it converges except when 1 or 3 parameters are used to model the lens distortions. Next, when no radial distortion was present the results also show that the best error buildup was obtained if no radial distortion was modeled. However, modeling it anyway the results did not completely deteriorate. Another important conclusion is that two radial parameters are often sufficient for convergence as the higher order terms only have a small contribution but using a higher number, e.g. 6, does not necessarily introduce divergence. Actually, we suspect a trade-off to exist. The exact ground truth radial distortion might be sufficiently modeled using only a few parameters. In this framework, however, the radial distortion needs to converge from an initial erroneous estimate to its final correct value. This convergence is a dynamic event in which more degrees of freedom increase the convergence rate. Once converged, a high number of parameters might lead to modeling the residual noise instead of the physical underlying lens distor-

Table 1. relative error buildup in pixels: frame 100x(i-1) - frame 100xi
Radial calibration matrix $\mathbf{K}_d$ with $f_d = 461.025$, $r_d = 1$, $s_d = 0$ and $(u_d, v_d) = (360, 288)$ on images of resolution 720×576. The number of radial parameters are counted excluding k_0.

Scenario	rad. param	i = 1	2	3	4	5	6	7	8	9
barrel	0	16.55	16.91	16.87	17.13	17.49	17.47	18.15	18.57	18.62
ground truth	1	12.09	12.87	13.87	14.83	15.66	16.95	18.71	20.76	22.19
	2	6.55	4.05	2.56	0.28	0.08	0.08	0.07	0.07	0.06
k_0 = 1.	3	26.73	46.34	54.46	87.63	130.52	197.76	242.24	79.49	35.18
k_1 = -0.01	4	6.86	5.67	5.85	5.42	3.52	3.57	4.67	5.06	7.45
k_2 = -0.01	6	3.04	0.90	0.12	0.11	0.08	0.06	0.09	0.07	0.07
k_3 = -0.01	8	5.03	4.58	3.79	1.30	0.18	0.08	0.08	0.08	0.08
	0	0.63	0.06	0.05	0.05	0.05	0.04	0.03	0.05	0.05
	1	5.92	8.59	10.52	12.09	13.39	13.06	10.64	10.21	8.93
	2	1.15	0.14	0.07	0.09	0.08	0.06	0.06	0.05	0.05
no radial	3	5.23	11.92	24.28	37.68	36.70	42.63	57.89	61.17	69.63
	4	1.18	0.30	0.08	0.08	0.08	0.06	0.09	0.11	0.09
	6	0.50	0.07	0.07	0.05	0.06	0.06	0.06	0.07	0.07
	8	3.86	1.63	0.08	0.07	0.07	0.08	0.06	0.07	0.06
pincushion	0	17.69	17.02	16.63	15.97	15.73	15.48	15.27	14.85	15.18
ground truth	1	15.92	15.24	14.88	15.29	16.38	19.41	21.24	20.63	19.41
	2	4.46	1.62	0.56	0.08	0.08	0.07	0.07	0.07	0.06
k_0 = 1.	3	13.33	21.10	34.93	45.30	42.28	47.66	51.76	54.16	56.45
k_1 = 0.01	4	3.05	0.79	0.10	0.09	0.08	0.08	0.06	0.07	0.07
k_2 = 0.01	6	3.30	4.60	2.56	1.03	0.14	0.12	0.09	0.08	0.07
k_3 = 0.01	8	5.00	4.88	6.16	6.11	6.30	6.05	6.07	7.18	6.31

tions. The strange phenomenon that odd number of parameters 1 and 3 do not converge still has to be investigated. Because of the complex interactions that are involved in this sequential structure, motion and lens distortion recovery a theoretical proof of convergence is very difficult. From the results, however, it is clear that not modeling present lens distortions always leads to bad results while better outcomes can be achieved by modeling the lens distortions. Figure 7 shows the difference in recovered camera positions of the turntable sequence with and without modeling of lens distortions. Without the consideration of distortions a systematic error is incorporated, clearly shown by the diverging camera path. When lens distortions are considered, they are captured and after an initial transient behavior the camera follows a periodic cyclic path as was the case in reality.

Figure 8 shows real-life footage (1200 frames, image resolution 720×576) representing the roof of an ancient fountain. The reconstruction without modeled lens distortions is shown in Figure 9. Clearly the roof which in reality is straight bends backwards in the reconstruction because of the present barrel distortion. When lens distortions were considered, modeling six radial parameters k_j, the reconstructions as shown in Figure 10 could be achieved. The first reconstruction in Figure 10 took two frames with zero radial distortion as a starting point. The second reconstruction took the final radial distortion values obtained by the

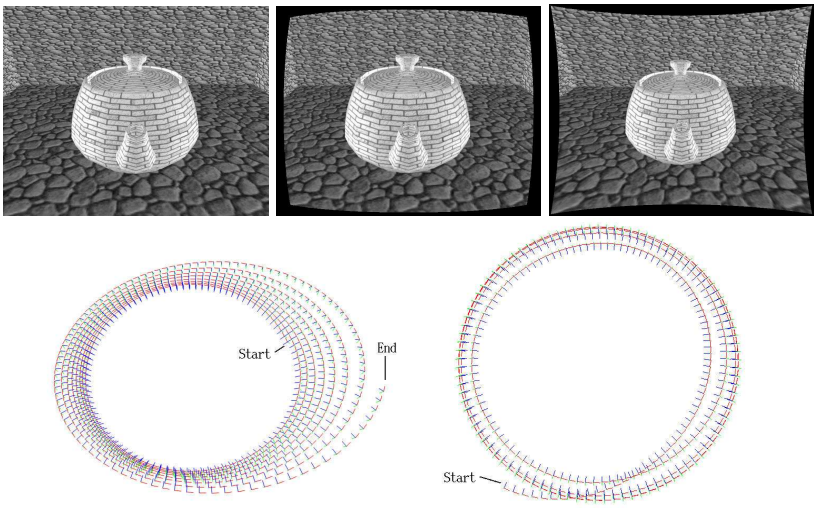

Fig. 7. Top: First frame of the original turntable sequence, the barrel distorted version and the pincushion distorted version respectively. Bottom Left: Divergence of the estimated camera path due to lack of radial distortion modeling when a barrel distorted version of the original turntable sequence was used. Bottom Right: Convergence of the camera path when the actual radial distortion is modeled. The camera path does not diverge but converges onto a circular path.

Table 2. relative error buildup in pixels: frame 100x(i-1) - frame 100xi

radial calibration matrix $\mathbf{K}_d$ with $f_d = 461.025$, $r_d = 1$, $s_d = 0$ and $(u_d, v_d) = (360, 288)$ on images of resolution 720×576. Variation of radial distortion parameters between $(k_0, k_1, k_2, k_3) = (1., -0.01, -0.01, -0.01)$ and $(k_0, k_1, k_2, k_3) = (1., -0.03, -0.03, -0.03)$

Scenario	rad. param	i = 1	2	3	4	5	6	7	8	9
barrel	6	3.05	0.83	0.08	0.07	0.08	0.08	0.08	0.07	0.08

previous run as a starting point and therefore could achieve a structure which closely resembles the real structure as no initial transients were present.

Because the strategy of sequentially modeling lens distortions allows for varying distortions, we tested another artificial video sequence similar to the teapot sequence. A round trip of the turntable sequence consists of 100 frames. We varied the radial distortion from one extreme at frame 0 to another extreme at frame 50, returning to the first extreme at frame 100. Again ten round trips were made and the error buildup assessed as shown in Table 2. Figure 11 shows how the estimated radial distortion moves between the two ground truth extreme radial distortions, proving that variable distortions pose no immediate problems.

6 Summary

In this paper we described how lens distortions could be recovered sequentially from a video sequence together with structure and motion without the need for preliminary distortion calibration or specific scene geometry. In a first section the camera model that approximates the real projection process was discussed. It consisted of two parts, an ideal pinhole camera model and a subsequent distortion model which could model radial and de-centering distortions. In a following section it was shown how the parameters of this camera model could be estimated in a multi-step iterative way given 3D-2D correspondences. The multi-step algorithm consisted of a separate estimation of the ideal pinhole camera model and the distortion parameters. This separation diminished the strong coupling between all the camera parameters which exists in a single global minimization formulation. The following sections explained how the pinhole camera parameters and the distortion parameters could be estimated by applying only linear least squares techniques. Subsequently, as the sequential nature of Structure, Motion and Distortion Recovery needs initialization it was shown that this consisted of 2 cameras with zero distortion or any off-line pre-calibration of the distortion if the video camera was still available. Experiments identified the advantages of taking lens distortions into account by demonstrating results on artificial and real video sequences. It showed that modeling lens distortion was crucial to the minimization of the error buildup. The negligence of lens distortions would introduce a systematic error which causes severe error buildup in recovered scene structure and cameras which could inhibit loop video sequences to close perfectly. Accurate camera retrieval enables the recognition and tracking of 3D scene points which were reconstructed at an earlier stage of the video processing. As demonstrated in [1] this recovery of scene points reduces drift in sequential algorithms by a large amount. This benefit would be lost if camera retrieval suffers from error buildup due to unmodeled lens distortions.

7 Future work

This work introduced the modeling of lens distortions in a sequential Structure and Motion Recovery framework. The lens distortions were modeled using the radial distortion formulation with a moving distortion center so that de-centering distortions could also be modeled. In [15] tangential and thin prism distortions are also modeled and these formulations could be used to even further optimize the camera model to best fit reality. However, too many free parameters could turn the process unstable and unreliable as one may be modeling noise instead of the physical projection process. This has to be investigated.

8 Acknowledgments

We would like to gratefully acknowledge the financial support of the FWO project G.0223.01 and the IST projects ATTEST and INVIEW.

References

1. K. Cornelis, M. Pollefeys, and L. Van Gool. Tracking based structure and motion recovery for augmented video productions. *Proceedings of the ACM Symposium on Virtual Reality Software and Technology VRST2001*, pages 17–24, November 2001.
2. J. Batista, H. Araújo, and A. Almeida. Iterative multi-step explicit camera calibration. *IEEE Transactions on Robotics and Automation*, 15(5), October 1999.
3. H. A. Beyer. Accurate calibration of ccd cameras. *In Proceedings of the International Conference on Computer Vision and Pattern Recognition, Urbana Champaign, Illinois, USA*, pages 96–101, June 1992.
4. F. Devernay and O. Faugeras. Automatic calibration and removal of distortion from scenes of structured environments. *In Proceedings of SPIE Conference, San Diego, CA*, 2567:62–72, July 1995.
5. Andrew W. Fitzgibbon. Simultaneous linear estimation of multiple view geometry and lens distortion. *CVPR, Kauai, Hawaii,*, I:125–132, December 2001.
6. R. Haralick, H. Joo, C. Lee, X. Zhuang, V. Vaidya, and M. Kim. Pose estimation from corresponding point data. *IEEE Transactions on Systems, Man and Cybernetics*, 19(6):1426–1446, November/December 1989.
7. E. Marchand and F. Chaumette. A new formulation for non-linear camera calibration using virtual visual servoing. *Rapport de Recherche IRISA, No 1366*, January 2001.
8. M. A. Penna. Camera calibration : A quick and easy way to determine the scale factor. *in IEEE Transactions on Pattern Analysis and Machine Intelligence*, 13(12):1240–1245, December 1991.
9. G. Stein. Internal camera calibration using rotation and geometric shapes. *Master's Thesis, Massachussets Institute of Technology. Artificial Intelligence Laboratory*, 1993.
10. G. Stein. Lens distortion calibration using point correspondences. *In Proceedings of the 1997 Conference on Computer Vision and Pattern Recognition*, pages 143–148, June 1997.
11. Stevenson, Daniel, and M. Fleck. Nonparametric correction of distortion. *TR 95-07, Comp. Sci., University of Iowa*, 1995.
12. R. Swaminathan and S. Nayar. Non-metric calibration of wide angle lenses. *In Proceedings of the 1998 DARPA Image Understanding Workshop, Monterey, California*, November 1998.
13. R. Y. Tsai. A versatile camera calibration technique for high accuracy 3d machine vision metrology using off-the-shelf tv cameras and lenses. *IEEE Journal of Robotics and Automation*, 3(4):323–344, August 1987.
14. G.-Q. Wei and S. Ma. Implicit and explicit camera calibration: Theory and experiments. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 16(5):469–480, May 1994.
15. J. Weng, P. Cohen, and M. Herniou. Camera calibration with distortion models and accuracy evaluation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 14(10):965–980, October 1992.
16. R. Willson. Modeling and Calibration of Automated Zoom Lenses. *Ph.D. thesis, Department of Electrical and Computer Engineering, Carnegie Mellon University*, January 1994.
17. Z. Zhang. On the epipolar geometry between two images with lens distortion. *In the Proceedings of the International Conference on Pattern Recognition (ICPR)*, I:407–411, August 1996.

Fig. 8. A couple of frames from a video sequence showing an ancient fountain roof. When pasted together, these give an idea of what the real roof looks like.

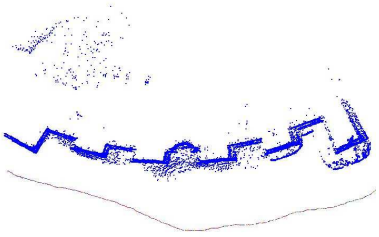

Fig. 9. Top view of the reconstruction of the fountain roof without modeling radial distortions. While the roof is straight in real life the curvature of the scene structure is clearly visible.

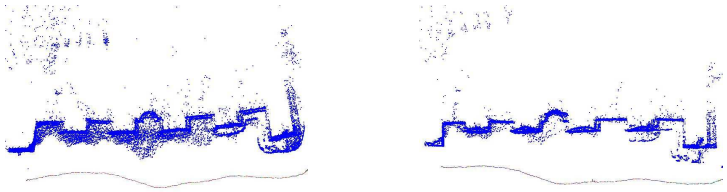

Fig. 10. Left: Top view of the reconstruction if the radial distortion was modeled with an initialization of 2 frames without radial distortion. Right: Result if the final radial distortion of a first run was taken as initialization for the program.

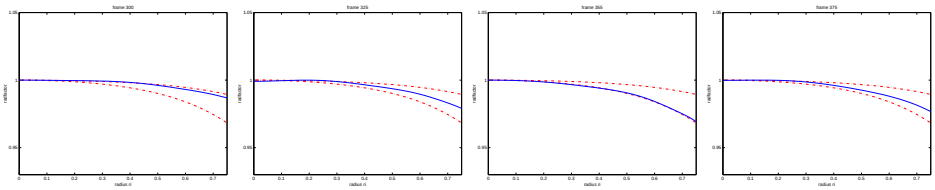

Fig. 11. From left to right [frames 300,325,355,375]: The variation of the estimated radial factor (solid curve) between the two ground truth extremes(dash-dot). The abscissa represents the radius in the ideal image plane Π_i. The ordinate represents the radial distortion factor *radfactor*.