3D Models from Extended Uncalibrated Video Sequences: Addressing Key-frame Selection and Projective Drift

Jason Repko
Department of Computer Science
University of North Carolina at Chapel Hill
repko@cs.unc.edu

Marc Pollefeys
Department of Computer Science
University of North Carolina at Chapel Hill
marc@cs.unc.edu

Abstract

In this paper, we present an approach that is able to reconstruct 3D models from extended video sequences captured with an uncalibrated hand-held camera. We focus on two specific issues: (1) key-frame selection, and (2) projective drift. Given a long video sequence it is often not practical to work with all video frames. In addition, to allow for effective outlier rejection and motion estimation it is necessary to have a sufficient baseline between frames. For this purpose, we propose a key-frame selection procedure based on a robust model selection criterion. Our approach guarantees that the camera motion can be estimated reliably by analyzing the feature correspondences between three consecutive views. Another problem for long uncalibrated video sequences is projective drift. Error accumulation leads to a non-projective distortion of the model. This causes the projective basis at the beginning and the end of the sequence to become inconsistent and leads to the failure of self-calibration. We propose a self-calibration approach that is insensitive to this global projective drift. After self-calibration triplets of key-frames are aligned using absolute orientation and hierarchically merged into a complete metric reconstruction. Next, we compute a detailed 3D surface model using stereo matching. The 3D model is textured using some of the frames.

1. Introduction

In recent years a lot of progress has been made in the area of 3D modeling from images. Several fully automatic systems have been proposed that reconstruct 3D models from video sequences recorded with a hand-held camera [10, 7, 6]. The development of statistically based algorithms have made it possible to automatically compute a robust estimate of the multi-view geometric relations between images of a sequence. In such cases, a projective 3D reconstruction can be obtained from an uncalibrated image sequence. The projective reconstruction can be upgraded to a metric reconstruction using a technique called self-calibration. Self-calibration achieves this by using constraints on the scene, camera motion, and camera intrinsics. Despite the substantial progress, many approaches that are effective on short sequences do not address the problems that can arise from processing extended sequences. In this paper we focus on two important problems that need to be addressed to reliably deal with extended uncalibrated video sequences.

When obtaining a 3D reconstruction from long image sequences, a subset of frames should be selected that are most suitable for the estimation of the epipolar geometry between views. The frames belonging to this subset are called *key-frames*. In order to accurately estimate the camera motion between frames, key-frames should be selected with a sufficient baseline between them. If two consecutive key-frames are too close to each other some of the degrees of freedom of the epipolar geometry are hard to estimate which leads to problems in motion estimation and in identifying divergent feature tracks. Different approaches have been proposed is in the literature [17, 3, 12]. All those approaches only perform a pairwise analysis for key-frame selection. However, as in the uncalibrated case, a triplet of views is the basic building block for structure and motion recovery, it is important to make sure that the key-frame selection takes this into account. A sufficient amount of well distributed features needs to observed by each set of three consecutive key-frames. This is not ensured by previous pairwise key-frame selection algorithms. The approach proposed in this paper is based on the model selection criterion proposed in [18]. The new key-frame selection criterion proposed in this paper uses features tracked over three frames.

A second issue that needs to be addressed when processing long image sequences is the mitigation of projective drift. Sequential structure and motion recovery is incrementally built-up by orienting additional views using the 3D positions of common features. Inaccuracies in the 3D positions of feature points due to noise and approximation errors accumulate over the course of the sequence and have an detrimental effect on the overall reconstruction. The estimated camera positions start to significantly deviate from their actual positions. Bundle adjustment, a common non-linear minimization technique used in refining the projective reconstruction, only partially helps with this problem. Steedly et al. [14], for example, have shown that for long video sequences significant deformations of the motion path often only result in small errors, or inversely that accurate motion can not be computed over extended video sequences without closing the loop. This projective drift is most evident in closed sequences where overlapping camera positions differ in the final reconstruction. However, this effect is just as important when the sequence is not closed. When radial distortion is not corrected before the structure and motion estimation, the drift can be much worse [2]. Projective drift is most problematic for self-calibration methods over long sequences, but has little effect over short sequences. Here we will propose a new self-calibration technique that is not sensitive to projective drift. The proposed approach is linear and imposes prior knowledge on the camera intrinsics and forces them to be constant over the whole sequence.

1.1. Notation

Points are represented by homogeneous 4-vectors $\mathtt{X}$ in 3-space, and by homogeneous 3-vectors $\mathtt{x}$ in the image. A plane is represented by a homogeneous 4-vector Π and a point $\mathtt{X}$ is on a plane if $\Pi^\top \mathtt{X} = 0$. A point $\mathtt{X}$ is mapped to its image $\mathtt{x}$ through perspective projection, represented by a 3×4 projection matrix $\mathbf{P}$ as $\lambda\mathtt{x} = \mathbf{P}\mathtt{X}$. The λ indicates a non-zero scale factor. In a metric coordinate system the matrix $\mathbf{P}$ can be factorized in intrinsic and extrinsic camera parameters: $\mathbf{P} = \mathbf{K}[\mathbf{R}\,\mathtt{t}]$ where the upper-triangular matrix $\mathbf{K}$ is given by the following equation:

$$\mathbf{K} = \begin{bmatrix} f & s & u \\ & rf & v \\ & & 1 \end{bmatrix} \tag{1}$$

with f the focal length (measured in pixels), r the aspect ratio, (u, v) the coordinates of the principal point and s a factor that is 0 when the pixels are rectangular. To deal with radial distortion, the perspective projection model is extended to $\mathbf{K}\mathcal{R}([\mathbf{R}\,\mathtt{t}]\mathtt{X})$ with $\mathcal{R}([x\,y\,1]^\top) = [x\,y\,w]^\top$, $w^{-1} = (1 + k_1 r^2 + k_2 r^4)$, $r^2 = x^2 + y^2$, and k_1 and k_2 are parameters of radial distortion. Two corresponding points $\mathtt{x}$ and $\mathtt{x}'$ should satisfy the epipolar constraint $\mathtt{x}'^\top \mathbf{F}\mathtt{x} = 0$. A point $\mathtt{x}$ located in the plane corresponding to the homography $\mathbf{H}$ is transferred from one image to the other according to $\lambda\mathtt{x}' = \mathbf{H}\mathtt{x}$. The fundamental matrix $\mathbf{F}$ and the two-image homography $\mathbf{H}$, are both 3×3 homogeneous matrices. A more complete description of these concepts can be found in [4].

2. Background

2.1. Uncalibrated structure and motion recovery

Given a set of corresponding image features, structure and motion recovery attempts to determine the metric 3D reconstruction of image features without any prior knowledge of the camera's calibration. Features are tracked between frames over a sequence of images. The epipolar geometry between frames can be determined from the feature correspondences, used to constrain the search for additional correspondences, and subsequently refined. The epipolar geometry is represented by a 3×3 matrix known as the fundamental matrix $\mathbf{F}$. If all features between frames are found on a planar object or the baseline between frames is not large enough, a planar homography can also describe the relation between frames. Consequently, the epipolar geometry can not be uniquely determined in some cases.

The projective reconstruction of the scene can be obtained from the epipolar geometry and the tracked features. Structure and motion recovery starts by setting up an initial projective reconstruction frame from the first two views and adding additional views using correspondences. The 3D structure is only known up to a projective transformation. Consequently, a 3D point's reprojection into the original image is used to determine and refine its position. Photogrammetric bundle adjustment is used to refine the projective reconstruction using the reprojection error of the feature's 3D point.

Projective reconstruction can be upgraded to a metric reconstruction using self-calibration. The Absolute Conic (AC) is a geometric entity determined by the feature correspondences that has a constant position relative to a moving camera. This is due to the fact that the AC is invariant under Euclidean transformations. Self-calibration locates and uses the AC to upgrade the reconstruction to metric. Inaccuracies in the estimations of the 3D point positions can cause the distortions in the projective model in turn affect location of the AC. The transformation that upgrades the reconstruction to metric can be solved for using the AC and the projection matrix, as discussed later.

2.2. Dense surface estimation and model creation

The reconstruction obtained as described in the previous section only contains a sparse set of 3D points. Therefore, the next step attempts to match all image pixels of an image with pixels in neighboring images, so that these points can also be reconstructed. This task is greatly facilitated by the knowledge of all the camera parameters that were obtained in the previous stage.

Since a pixel in the image corresponds to a ray in space and the projection of this ray in other images can be predicted from the recovered pose and calibration, the search of a corresponding pixel in other images can be restricted to a single line. Additional constraints, such as the assumption of a piecewise continuous 3D surface, are also employed to further constrain the search. It is possible to warp the images so that the search range coincides with the horizontal scanlines. This allows us to use an efficient stereo algorithm that computes an optimal match for the whole scanline at once. Thus, we can obtain a depth estimate (i.e. the distance from the camera to the object surface) for almost every pixel of an image.

A complete dense 3D surface model is obtained by fusing the results of all the depth images together. The images used for the reconstruction can also be used for texture mapping so that a final photo-realistic result is achieved. The different steps of the process are illustrated in Figure 1. An in-depth description of this approach can be found in [11].

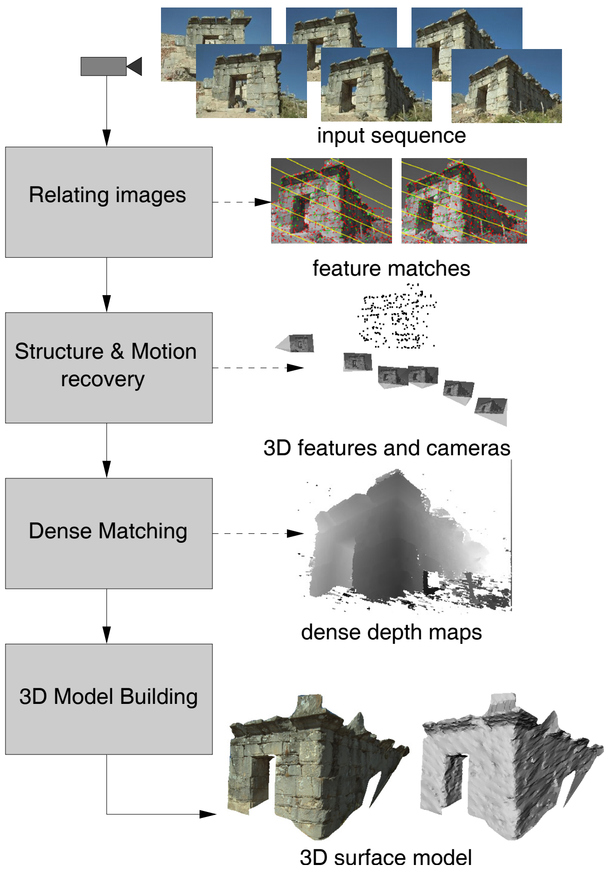

Figure 1. Overview

3. Solving the key-frame selection problem

Our key-frame criterion selects views that are best suited for computing the three-view geometry. Each view is evaluated based on features that it has in common with the previous two key-frames. The remainder of the structure and motion recovery procedure uses triplets of key-frames with a two view overlap.

The Geometric Robust Information Criterion (GRIC) model selection approach is useful in key-frame selection [18]. The GRIC computes a score based on the feature correspondences between images and a model. Two models can be quantitatively compared to each other because scores are assigned to models based on the model's parsimony and how well the model fits the data. The model with the lowest score is a best fit.

$$GRIC = \sum \rho(e_i^2) + (nd \ln r + k \ln rn), \qquad (2)$$

such that,

$$\rho(e_i^2) = \min\left(\frac{e^2}{\sigma^2}, 2(r-d)\right) \qquad (3)$$

In the above equation, if the dimension of the data is r and n is the total number of features then $nd\ln(r)$ represents the penalty term for the structure having parameters that are n times the manifold's dimension d. Each parameter is estimated from r observations. The term $k\ln(rn)$ represents the penalty for the motion model having k parameters estimated from rn observations.

Between each pair of images, a Maximum Likelihood Estimator (MLE) is used to determine the fundamental matrix **F** and 2D homography **H** using the feature correspondences. Since **F** can describe the simpler model **H** because $\mathbf{F} = [\mathtt{e}']_x \mathbf{H}$ (with the epipole $\mathtt{e}'$ corresponding to the left nullspace of $\mathtt{F}$), the GRIC(**F**) will produce a lower score if **F** best fits the data. When no translation is observed, techniques that estimate **F** will fit the outliers to the solution, producing meaningless results. In addition, the **H** model best describes the correspondences for views with a small baseline. By using only views that are related by **F**, the epipolar geometry between two views is unambiguously recovered by avoiding pure rotation and small baselines. Furthermore, with an robust estimate of **F**, points that do not fit the epipolar constraints can be removed from the system. As a result, the projective structure can unambiguously be recovered in the standard way.

The key-frame criterion proposed selects three views where the GRIC score of the epipolar model is lower than the score of the homography model. In our algorithm, the frames are processed sequentially, and the first view in the sequence is selected as the first key-frame. The second key-frame is determined by the last view to share more than 90% of the features with the view that GRIC(**F**) was smaller than GRIC(**H**) as in [13]. Subsequent frames are selected as key-frames based on the GRIC with features that are found not only in the last key-frame selected but also in the second to last key-frame. Thus, a view is selected as the a key-frame if the fundamental matrix between it and the last key-frame view yields a GRIC(**F**) score that is smaller than the

GRIC($\mathbf{H}$) score of the homography between the two views. The features used in computing the GRIC are those from the current view that are found in the previous two key-frames. By using features found in the current view and the previous two key-frame views, we ensure that the three-view geometry can be estimated reliably for each selected triplet of views.

Using points tracked over three views also is an indication that the tracked points are well supported. The probability of a tracked feature point being an outlier is diminished. Each view that the feature is tracked in will contribute evidence supporting its 3D position. Our objective is to obtain a strong local structure and motion solution for each triplet. As each key-frame is selected, the triplet of key-frames (the current selection along with the previous two) are reconstructed, and each reconstruction is refined using bundle adjustment. Each projectively reconstructed triplet is separate from the others. As we will see in the next section, the projective drift problem is made less severe by working with triplets of views. A global structure and motion solution will be obtained by merging triplets after self-calibration.

4. Solving the projective drift problem in long sequences

Once the projective structure and motion has been computed, self-calibration can be used to restrict the ambiguity on the reconstruction from a projective to a similarity transformation (i.e. a Euclidean transformation plus scaling). However, for long image sequences self-calibration techniques that require one single consistent projective reconstruction, e.g. [19, 9], tend to fail. The reason for this is the error accumulated during the structure and motion recovery. The result is that self-calibration constraints at the beginning and at the end of the sequence might end up not being consistent. In other words, the solution for the absolute conic that one would obtain from the first part of the sequence, would not satisfy the constraints on the intrinsic camera parameters for the last part and vice-versa. We have regularly observed this with real sequences where self-calibration would fail for long sequences while it would be successful for a subsequence.

One possible solution to this problem is to use a self-calibration algorithm that does not enforce a global consistency and only directly uses information from nearby images. The Kruppa equations [1] could for example be used since they only use pairwise information, but they suffer from additional degeneracies [16]. Also, almost all successful algorithms based on the Kruppa equations use fundamental matrices between remote views in the sequence derived from the projective reconstruction. In this case the Kruppa equations are also sensitive to error accumulation in the reconstruction process.

Here we propose an alternative algorithm that is not sensitive to error accumulation on the projective reconstruction. Our approach is based on the coupled self-calibration presented in [13] to deal with degeneracies for structure and motion recovery in the presence of dominant planes. We will only require projective consistency of the reconstruction for triplets of neighboring views, but at the same time force the intrinsic camera parameters for the whole sequence to be constant. The proposed approach is linear. The approach is based on the projection equation for the absolute quadric [19]:

$$\lambda \mathbf{K}\mathbf{K}^\top = \mathbf{P}\mathbf{\Omega}^*\mathbf{P}^\top \tag{4}$$

where $\mathbf{\Omega}^*$ represents the absolute quadric. In metric space $\mathbf{\Omega}^* = \mathrm{diag}(1,1,1,0)$, in projective space $\mathbf{\Omega}^*$ is a 4×4 symmetric rank 3 matrix representing an imaginary disc-quadric. By transforming the image so that a typical focal length (e.g. 50mm) corresponds to unit length in the image and that the center of the image is located at the origin, realistic expectations for the intrinsics are $\log(f) = \log(1) \pm \log(3)$ (i.e. f is typically in the range $[17\text{mm}, 150\text{mm}]$), $r = log(1) \pm \log(1.1), u = 0 \pm 0.1, v = 0 \pm 0.1, s = 0$. These expectations can be used to obtain a set of weighted self-calibration equations from Equation (4):

$$\begin{array}{rcl}
\frac{1}{9\lambda}\left(P_1\mathbf{\Omega}^*{P_1}^\top - P_3\mathbf{\Omega}^*{P_3}^\top\right) & = & 0 \\
\frac{1}{9\lambda}\left(P_2\mathbf{\Omega}^*{P_2}^\top - P_3\mathbf{\Omega}^*{P_3}^\top\right) & = & 0 \\
\frac{1}{0.2\lambda}\left(P_1\mathbf{\Omega}^*{P_1}^\top - P_2\mathbf{\Omega}^*{P_2}^\top\right) & = & 0 \\
\frac{1}{0.01\lambda}\left(P_1\mathbf{\Omega}^*{P_2}^\top\right) & = & 0 \\
\frac{1}{0.1\lambda}\left(P_1\mathbf{\Omega}^*{P_3}^\top\right) & = & 0 \\
\frac{1}{0.1\lambda}\left(P_2\mathbf{\Omega}^*{P_3}^\top\right) & = & 0
\end{array} \tag{5}$$

where P_i is the i-th row of a projection matrix and λ a scale factor that is initially set to 1 and later on to $P_3\tilde{\mathbf{\Omega}}^*{P_3}^\top$ with $\tilde{\mathbf{\Omega}}^*$ the result of the previous iteration. In practice iterating is not really necessary, but a few iterations can be performed to refine the initial result. Experimental validation has shown that this approach yields much better results than the original approach described in [9]. This is mostly due to the fact that constraining all parameters (even with a small weight) allows to avoid most of the problems due to critical motion sequences [15] (especially the specific additional case for the linear algorithm [8]).

When choosing $\mathbf{P} = [\mathbf{I}|\mathbf{0}]$ for one of the projection matrices it can be seen from Equation (4) that $\mathbf{\Omega}^*$ can be written as:

$$\mathbf{\Omega}^* = \left[\begin{array}{cc} \mathbf{K}\mathbf{K}^\top & \mathbf{a} \\ \mathbf{a}^\top & b \end{array}\right] \tag{6}$$

Now the set of equations (5) can thus be written as:

$$\begin{bmatrix} \mathbf{C} & \mathbf{D} \end{bmatrix} \begin{bmatrix} \mathbf{k} \\ \begin{bmatrix} \mathbf{a} \\ b \end{bmatrix} \end{bmatrix} \tag{7}$$

where $\mathbf{k}$ is a vector containing six coefficients representing the matrix $\mathbf{KK}^\top$, $\mathbf{a}$ is a 3-vector and b a scalar and $\mathbf{C}$ and $\mathbf{D}$ are matrices containing the coefficients of the equations. We propose to write down those equations for each triplet of consecutive views.

If the sequence is recorded with constant intrinsics, the vector $\mathbf{k}$ will be common to all triplets and one obtains the following coupled self-calibration equations:

$$\begin{bmatrix} \mathbf{C}_1 & \mathbf{D}_1 & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{C}_2 & \mathbf{0} & \mathbf{D}_2 & \cdots & \mathbf{0} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \mathbf{C}_n & \mathbf{0} & \mathbf{0} & \cdots & \mathbf{D}_n \end{bmatrix} \begin{bmatrix} \mathbf{k} \\ \begin{bmatrix} \mathbf{a}_1 \\ b_1 \end{bmatrix} \\ \begin{bmatrix} \mathbf{a}_2 \\ b_2 \end{bmatrix} \\ \vdots \\ \begin{bmatrix} \mathbf{a}_n \\ b_n \end{bmatrix} \end{bmatrix} \tag{8}$$

The advantage of this approach is that it is far less sensitive to projective drift since we only require consistency between triplets of consecutive views, while other approaches require consistency over the whole sequence. For each subsequence a transformation to upgrade the reconstruction from projective to metric can be obtained from the constraint $\mathbf{T}_i \mathbf{\Omega}_i^* \mathbf{T}_i^\top = \mathrm{diag}(1,1,1,0)$ (through eigenvalue decomposition).

To validate our approach to deal with projective drift we have generated a synthetic 30 view sequence observing 100 points. The camera motion consists of a 180 degree rotation around the center of the points. For each view we generate a projective transformation close to identity and apply it to that view and all the ones that follow to simulate projective drift. Then, we apply both the standard self-calibration algorithm on the 30 view sequence and the triplet-based self-calibration algorithm to attempt to upgrade the reconstruction from projective to metric. Both results can be seen in Figure 2. The standard approach fails (as the circular path gets warped to a hyperbola), while the proposed approach yields good results. Notice that the small deviation from the original half circle that is visible for the successfully upgraded reconstruction corresponds to the induced projective drift.

4.1. Merging the reconstructions

To merge the subsequences into a global reconstruction, corresponding 3D features and camera positions of adjacent triplets are registered using the absolute orientation [5]. Absolute orientation is a least-squares solution that provides the Euclidean transformation between the 3D to 3D feature correspondences between the two overlapping views of the adjacent subsequences. Ideally, after self-calibration each separate metric reconstruction should differ only by a Euclidean transformation. However, inconsistencies in the 3D to 3D correspondences due to inaccurate estimations of 3D positions may exist. Absolute orientation must take into account the presence of outliers in the 3D to 3D correspondences. Absolute orientation followed by bundle adjustment is performed on adjacent subsequences hierarchically. Metric bundle adjustment is used to further refine the aligned subsequences by enforcing the alignment using redundant 3D points and cameras. These redundant points and cameras are removed from the reconstruction before a final bundle adjustment.

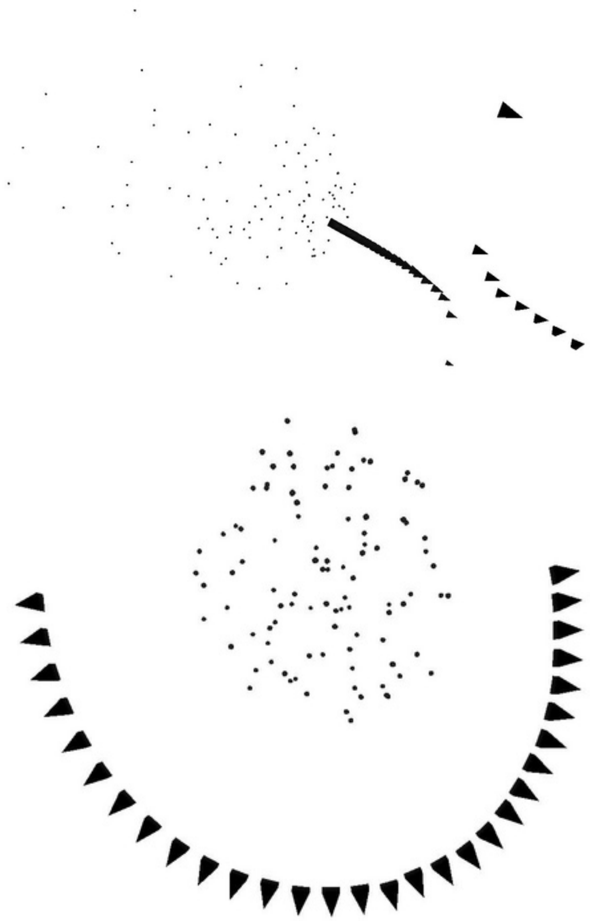

Figure 2. Failure of self-calibration due to projective drift (top), success of proposed approach in dealing with projective drift (bottom)

5. Results And Discussion

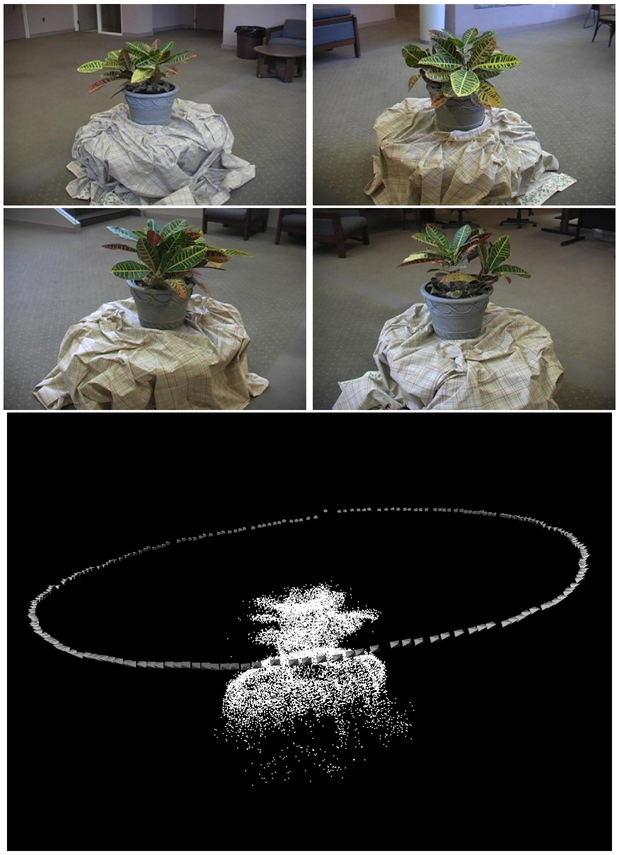

Figure 3. Our algorithm selected 204 key-frames from 983 frames of video of a potted plant. Four key-frames are shown along with the sparse 3D point structure and recovered camera positions.

We have used the reconstruction algorithm on multiple sequences of varying length. A few examples are shown here. The result of using our reconstruction algorithm on a 983 frame video sequence of a potted plant is shown in Figure 3. Our algorithm selected 204 key-frames.

To illustrate how the GRIC values are related to the baseline, another video sequence was created by walking around a potted plant and pausing at each step while continually filming. This camera motion is reflected in the recovered camera positions as shown in Figure 6. In Figure 7, the HGRIC and FGRIC scores alternate for subsequences corresponding to periods where the camera is in motion, and the HGRIC score remains lower than the FGRIC score for the subsequences corresponding to periods where the camera is relatively static. The alternation between the HGRIC and FGRIC scores is due to a key-frame being selected and then compared to the frames that immediately follow it. Once a sufficient baseline is established between the previous key-frame and the frame being examined, the FGRIC score drops below the HGRIC score and the corresponding frame becomes a key-frame.

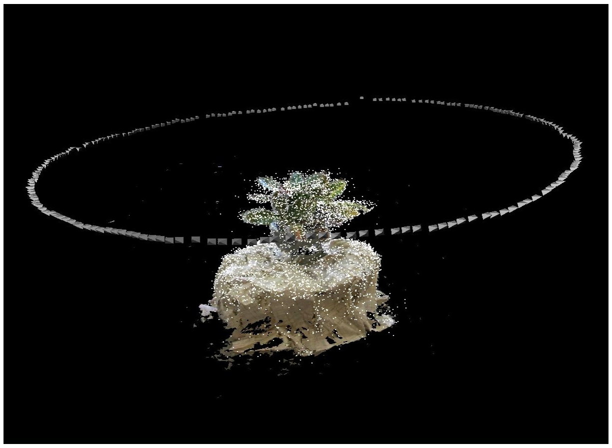

Figure 4. A model constructed from the potted plant sequence showing the sparse 3D point structure and recovered camera positions.

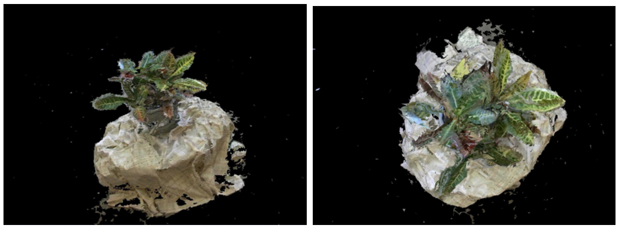

Figure 5. Side and top views of the model constructed from potted plant sequence.

For the video sequence shown in Figure 8, the recovered camera positions appear to follow a circular path, and the last view appears to be close to the initial. No attempt was made to enforce the sequence to be closed. The reconstruction indicates that the recovered camera position for last the key-frame chosen was relatively close to the first. If significant projective drift was present then the first and last cameras would not appear as close as shown in the reconstruction in Figure 8.

Noisy point features may be a result of features that are tracked, lost, and subsequently come back into view. There is no information linking these points together, and as a result their 3D positions may slightly drift apart. The large amount of features is a result of our triplet based approach. In addition, the construction is tolerant of subsequences that have a residual error from bundle adjustment that is above

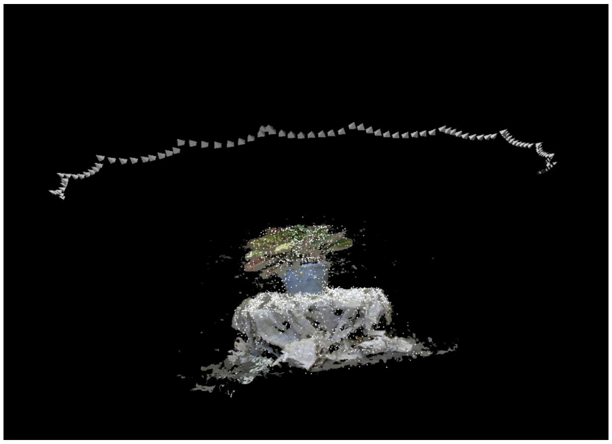

Figure 6. The camera path reflects the exaggerated walking motion of the camera operator. Each arc in the camera path corresponds to a step. The same scene as shown in Figure 3 is used.

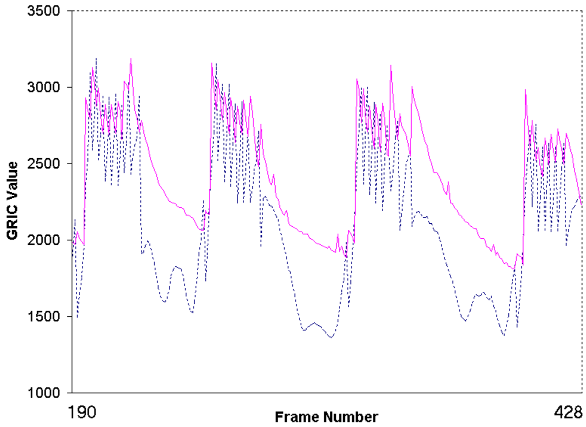

Figure 7. A graph of the FGRIC (shown as a solid line) and HGRIC (shown as a dashed line) as they vary over a portion (238 frames) of the video sequence. The score for each frame is computed using the fundamental matrix $\mathbf{F}$ or 2D homography $\mathbf{H}$ determined by the feature correspondences found in the frame and the previous two key-frames.

the sequence's mean if the surrounding subsequences have a low residual. Typically, the error will be distributed over the adjacent subsequences.

In some cases, after self-calibration, subsequences may still have a small amount of distortion, which introduces error into the result of the absolute orientation. Consequently, the cameras and points may not be in exact alignment. In our experiments, we have found that the metric bundle following the absolute orientation calculation reduces the effects of a poor alignment.

6. Conclusions And Future Work

This paper has addressed two issues associated with processing long video sequences: key-frame selection and projective drift. A new key-frame selection criterion based on the GRIC was presented that uses features tracked over three views. Our three view approach allows us to estimate the projective structure using well supported features. As demonstrated, when sequentially updating a reconstruction, projective drift causes cameras to significantly deviate from their actual position. Using triplets of key-frames with coupled self calibration avoids this by maintaining separate local reconstructions and enforcing constant camera intrinsics over all of the key-frame subsequences.

The system proposed in this paper can be enhanced by handling the case where all feature tracks are lost between views during key-frame selection. In the case where all points are found on dominant planar features over a long sequence of frames, H will be the best model. In this case, the number of feature points may decrease significantly as the baseline between the current frame and the last key-frame increases. Our results show that this is not a significantly limiting case, but it is an interesting avenue for future work.

Acknowledgements

The authors would like to acknowledge the financial support of the NSF grant IIS-0313047. Part of the software presented in this paper was developed at the K.U.Leuven.

References

[1] O. Faugeras. What can be seen in three dimensions with an uncalibrated stereo rig? In *Computer Vision – ECCV'92, LNCS, Vol.588*, pages 563–578. Springer-Verlag, 1992.

[2] A. Fitzgibbon. Simultaneous linear estimation of multiple view geometry and lens distortion. In *Conference on Computer Vision and Pattern Recognition (1)*, pages 125–132, 2001.

[3] S. Gibson, J. Cook, T. Howard, R. Hubbold, and D. Oram. Accurate camera calibration for off-line, video-based augmented reality. In *IEEE and ACM International Symposium on Mixed and Augmented Reality (ISMAR)*, 2002.

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

[5] B. K. P. Horn, H. M. Hilden, and S. Negahdaripour. Closed-form solution of absolute orientation using orthonormal matrices. *Journal of the Optical Society of America. A, Optics and image science*, 5(7):1127–1135, July 1988.

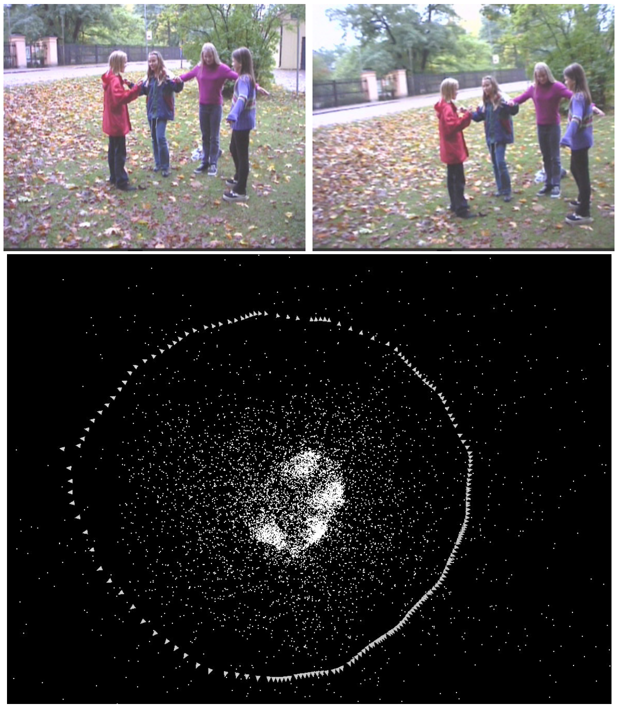

Figure 8. Our algorithm selected 176 key-frame views from 541 frames. The reconstruction above is accompanied by the first and last key-frames from the video sequence. The first (left) and last (right) key-frames from the original image sequence appear to be taken at close positions, which is corroborated by our reconstruction. Image sequence courtesy of David Nister.

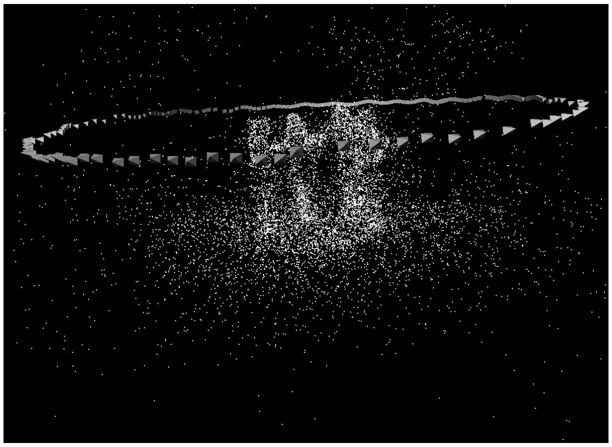

Figure 9. A side view of the reconstruction.

[6] M. Lhuillier and L. Quan. A quasi-dense approach to surface reconstruction from uncalibrated images. *to appear in IEEE Transaction on Pattern Analysis and Machien Intelligence*.

[7] D. Nister. *Automatic dense reconstruction from uncalibrated video sequences*. PhD thesis, Royal Institute of Technology KTH, Stockholm, Sweden, 2001.

[8] M. Pollefeys. *Self-calibration and metric 3D reconstruction from uncalibrated image sequences*. PhD thesis, Dept. of Electrical Engineering, Katholieke Universiteit Leuven, 1999.

[9] M. Pollefeys, R. Koch, and L. Van Gool. Self-calibration and metric reconstruction in spite of varying and unknown internal camera parameters. In *Proc. ICCV'98*, pages 90–95, 1998.

[10] M. Pollefeys, R. Koch, and L. Van Gool. A simple and efficient rectification method for general motion. In *Proc.ICCV'99*, pages 496–501, 1999.

[11] M. Pollefeys, L. Van Gool, Vergauwen M., Verbiest F., Cornelis K., Tops J., and Koch R. Visual modeling with a hand-held camera. *International Journal of Computer Vision*, 59(3):207–232, 2004.

[12] M. Pollefeys, L. Van Gool, M. Vergauwen, K. Cornelis, F. Verbiest, and J. Tops. Video-to-3d. In *Proceedings of Photogrammetric Computer Vision 2002 (ISPRS Commission III Symposium), International Archive of Photogrammetry and Remote Sensing. Volume 34*, pages 252–258, 2002.

[13] M. Pollefeys, F. Verbiest, and L. Van Gool. Surviving dominant planes in uncalibrated structure and motion recovery. In *Computer Vision - ECCV 2002, LNCS, Vol.2351*, pages 837–851, 2002.

[14] D. Steedly, I. Essa, and Dellaert F. Spectral partitioning for structure from motion. In *International Conference on Computer Vision*, 2003.

[15] P. Sturm. Critical motion sequences for monocular self-calibration and uncalibrated euclidean reconstruction. In *Proc. CVPR'97*, pages 1100–1105, 1997.

[16] P. Sturm. A case against kruppa's equations for camera self-calibration. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(10):1199–1204, Sep 2000.

[17] T. Thormählen, H. Broszio, and A. Weissenfeld. Keyframe selection for camera motion and structure estimation from multiple views. In *Computer Vision – ECCV 2004, LNCS, Vol. 3021*, pages 523–535, 2004.

[18] P. Torr, A. Fitzgibbon, and A. Zisserman. Maintaining multiple motion model hypotheses through many views to recover matching and structure. In *Proc. ICCV'99*, pages 485–491, 1998.

[19] B. Triggs. The absolute quadric. In *Proc. CVPR'97*, pages 609–614, 1997.