# Flexible 3D Acquisition with a Monocular Camera

Marc Pollefeys, Reinhard Koch, Maarten Vergauwen and Luc Van Gool
ESAT–VISICS, K.U.Leuven,
Kard.Mercierlaan 94, B-3001 Heverlee, Belgium
{firstname.lastname}@esat.kuleuven.ac.be

*submitted to Int'l Conf. on Robotics and Automation'98*

## Abstract

*One of the key problems for autonomous vehicles is the acquisition of 3D information about their environment. In this paper a flexible technique for 3D acquisition is proposed. This technique only requires an uncalibrated monocular camera. No prior knowledge about the scene or about the camera is necessary to build metric 3D models. In addition zoom and focus can be used freely. The feasability of the approach has been tested on both real and synthetic data and is illustrated here on real image sequences.*

## 1 Introduction

In the last few years the interest in 3D models has dramatically increased. More and more applications are using computer generated models. The main difficulty lies with the model acquisition. Although more tools become available to ease the generation of models, it is still a time consuming and expensive process. In many cases models of existing scenes or objects are desired. Traditional solutions include the use of stereo rigs, laser range scanners and other 3D digitizing devices. These devices are often very expensive, require careful handling and complex calibration procedures and are designed for a restricted depth range only.

In this work an alternative approach is proposed which avoids most of the problems mentioned above. The scene which has to be modeled is recorded from different viewpoints by a video camera. The relative position and orientation of the camera and its calibration parameters will automatically be retrieved from image data. Hence, there is no need for measurements in the scene or calibration procedures whatsoever. There is also no restriction on range, it is just as easy to model a small object, as to model a complete building. This will be shown in the examples. The proposed method thus offers a previously unknown flexibility in 3D model acquisition. In addition, no more than a camcorder is needed for scene acquisition. Hence, increased flexibility is accompanied by a decrease in cost.

This flexibility opens the way to new applications. Scenes filmed with a simple hand-held camcorder can be reconstructed. Models can be generated from old film footage (e.g. from monuments destroyed during the war). It will become possible to generate realistic 3D models of complete sites (e.g. archeological sites). Besides this, 3D modeling of objects (e.g. for tele-shopping applications or virtual exhibitions) also becomes a lot easier.

## 2 Model Acquisition

Two things are needed to build a 3D model from a image sequence: (1) the calibration of the camera setup[1] and (2) the correspondences between the images. Starting from an images sequence acquired by an uncalibrated video camera, both these prerequisites are unknown and therefore have to be retrieved from image data. At least a few correspondences are needed to retrieve the calibration of the camera setup, but this calibration facilitates the search for correspondences a lot.

### 2.1 Retrieving the Projective Framework

The first correspondences are found by extracting intensity corners in different images and matching them using a robust algorithm [9]. In conjunction with the matching of the corners the projective calibration of the setup is calculated. This allows to eliminate matches which are inconsistent with the calibration. Using the

[1] By *calibration* we mean the actual internal calibration of the camera as well as the relative position and orientation of the camera for the different views.

**Figure 1. (a) a priori search region, (b) search region based on epipolar geometry , (c) search region after projective reconstruction (used for refinement).**

projective calibration more matches can easily be found and used to refine this calibration. This can be seen in Figure 1.

The matching is first carried out on the first two images. This defines a projective framework in which the projection matrices of the other views are retrieved one by one. In this approach we follow the procedure proposed by Beardsley *et al* [1]. We therefore obtain projection matrices ($3 \times 4$) of the following form:

$$\mathbf{P}_1 = [\mathbf{I}|0] \text{ and } \mathbf{P}_k = [\mathbf{H}_{1k}|e_{1k}] \tag{1}$$

with $\mathbf{H}_{1k}$ the homography for some reference plane from view 1 to view $k$ and $e_{1k}$ the corresponding epipole.

## 2.2 Retrieving the Metric Framework

Such a projective calibration is certainly not satisfactory for the purpose of 3D modeling. A reconstruction obtained up to a projective transformation can differ very much from the original scene according to human perception: orthogonality and parallellism are in general not preserved, part of the scene can be warped to infinity, etc. To obtain a better calibration, constraints can be formulated by assuming that the internal camera parameters stay constant. By exploiting these constraints, the projective reconstruction can be upgraded to affine or even to metric (Euclidean up to scale).

Assuming constant camera parameters means that the projection matrices should all be of the following form:

$$\mathbf{P}_{Ek} = \mathbf{K}[\mathbf{R}_{1k}|\text{-}\mathbf{R}_{1k}t_{1k}] \tag{2}$$

with $\mathbf{K}$ an upper triangular matrix containing the (constant) intrinsic camera parameters, $t_{1k}$ and $\mathbf{R}_{1k}$ indicating the position and orientation of the camera for view $k$. From this it follows that the homographies of the plane at infinity between two views have the following form: $\mathbf{H}_{\infty kl} = \mathbf{K}\mathbf{R}_{kl}\mathbf{K}^{-1}$ and are therefore conjugated with the rotation matrix $\mathbf{R}_{kl}$. This yields us a constraint which can be used to obtain the affine calibration [6, 7]. Once the affine calibration has been obtained it is easy to upgrade this to metric through the use of the following equations:

$$\mathbf{K}\mathbf{K}^\top = \mathbf{H}_{\infty kl}\mathbf{K}\mathbf{K}^\top\mathbf{H}_{\infty kl}^\top \ . \tag{3}$$

The calibration parameters can be obtained from $\mathbf{K}\mathbf{K}^\top$ by Cholesky factorization.

Other methods exist for self-calibration (see for example [8]). Some of these methods can even be extended to varying focal length (assuming that rows and columns are orthogonal and that the aspect ratio is known) or after some initialization involving a pure translation [5].

## 2.3 Retrieving Dense Correspondences

At this point we dispose of a sparse metric reconstruction. Only a few points are reconstructed. Obtaining a dense reconstruction could be achieved by interpolation, but in practice this does not yield satisfactory results. Often some salient features are missed during the corner matching and will therefore not appear in the reconstruction. If for example the corner of the roof is missing, this could result in a whole part of the roof missing when using interpolation.

These problems can be avoided by using algorithms which estimate correspondences for almost every point in the images. At this point algorithms can be used which were developed for calibrated 3D systems like stereo rigs. Since we have computed the projective calibration between successive image pairs we can exploit the epipolar constraint that restricts the correspondence search to a 1-d search range. In particular it is possible to remap the image pair to standard geometry with the epipolar lines coinciding with the image scan lines [3]. The correspondence search is then reduced to a matching of the image points along each image scanline. In addition to the epipolar geometry other constraints like preserving the order of neighboring pixels, bidirectional uniqueness of the match, and detection of occlusions can be exploited. These constraints are used to guide the correspondence towards the most probable scanline match using a dynamic programming scheme [2].

### 2.4 Building the Model

Once a dense correspondence map and the metric camera parameters have been estimated, dense surface depth maps are computed using depth triangulation. The 3D model surface is constructed of triangular surface patches with the vertices storing the surface geometry and the faces holding the projected image color in texture maps. The texture maps add very much to the visual appearance of the models and augment missing surface detail.

The model building process is at present restricted to partial models computed from image pairs from single view points and work remains to be done to fuse different view points. Since we work in a metric framework it is possible to fuse the depth estimate from the view points within the framework into one consistent model surface [3].

Sometimes it is not possible to obtain a single metric framework for large objects like buildings since one may not be able to record images continuously around it. In that case the different frameworks have to be registered to each other. This will be done using available surface registration schemes [4].

## 3 Experiments

The feasability of the approach has been verified on synthetic data as well as on real video sequences. In Figure 2 some images of a video sequence of 24 images are shown which were used to obtain a 3D model of a part of the Arenberg castle in Leuven. The sequence was taken by freely moving a hand held camcorder. The complete sequence was used to obtain the calibration, while only two images were used for dense

**Figure 2. Images of the Arenberg castle which were used to generate the 3D model shown in Figures 3.**

**Figure 3. Perspective views of the 3D reconstruction obtained from the sequence seen in Figure 2.**

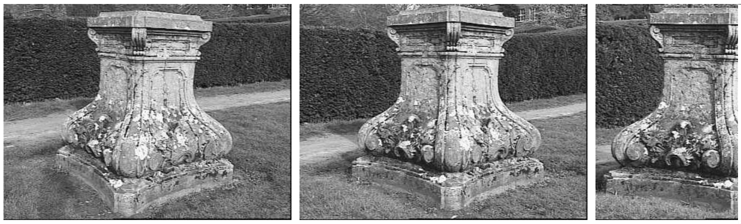

**Figure 4. Sequence 2**

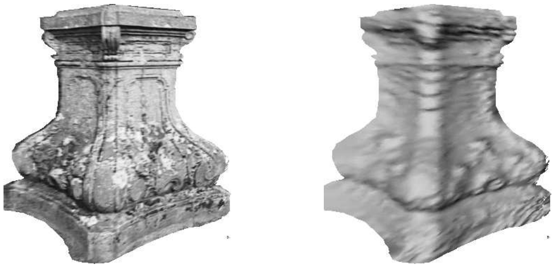

**Figure 5. Perspective view of the reconstruction (with texture and with shading).**

reconstruction. The 3D reconstruction is stored as a textured wire-frame VRML model. A few perspective views of this model can be seen in Figure 3. A more quantitative evaluation was obtained by measuring angles in the reconstructed scene between parallel lines (1.8 ± 1.1 degrees) and orthogonal lines (89.7 ± 1.4 degrees). These results confirm the good metric calibration obtained by the method.

As a second example 8 images of a stone pillar with curved surfaces were taken. Figure 4 show 3 of the recorded images. While filming and moving away from the object the zoom was changed (2×) to keep the image size of the object constant[2]. In spite of the changes in focal length the metric frame could be retrieved through self-calibration.

In Figure 5 a perspective view of the reconstruction is given, rendered both shaded and with surface texture mapping. The shaded view shows that even most of the small details of the object are retrieved.

Figure 6 shows three perspective views of the reconstructed object. Although there is some distortion at the outer boundary of the object, a highly realistic impression of the object is given. Note the arbitrarily shaped free-form surface that has been reconstructed.

A quantitative assessment of the metric properties for the pillar is not so easy because of the curved surfaces. It is, however, possible to measure some distances on the real object as reference lengths and compare them with the reconstructed model. In this case it is possible to obtain a measure for the absolute scale and verify the consistency of the reconstructed length within the model. Averaging all measured distances gave a consistant scale factor of 40.25 with a standard deviation of 5.4% overall. For the interior distances (avoiding the inaccuracies at the border of the model), the reconstruction error dropped to 2.3%. These results demonstrate the metric quality of the reconstruction even for complicated surface shapes and varying focal length.

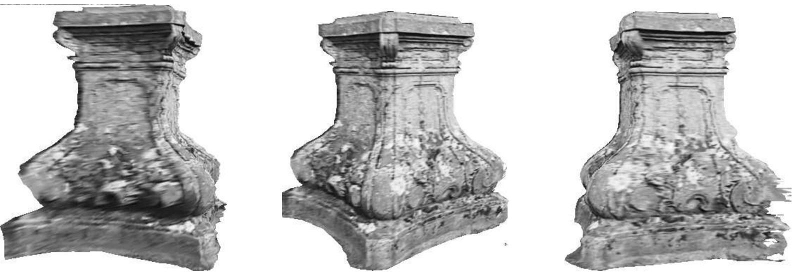

**Figure 6. Three perspective views of the reconstruction (left, front, right).**

## 4 Conclusion

An automatic 3D scene modelling technique was discussed that is capable of building models from uncalibrated image sequences. The technique is able to extract metric 3D models without any prior knowldege about the scene or the camera. The calibration is obtained by assuming a rigid scene and some constraints on the camera parameters (e.g. constant internal parameters).

Work remains to be done to get more complete models by integrating dense correspondences between more than 2 images. This will also increase the accuracy of the models and eliminate artefacts at the occluding boundaries. For this we can rely on work already done for calibrated systems.

## Acknowledgements

We would like to thank Andrew Zisserman and the rest of the VANGUARD team in Oxford for their contributions to the software used for this work. A specialization grant from the Flemish Institute for Scientific Research in Industry (IWT) and the financial support from the EU ACTS project AC074 'VANGUARD' are also gratefully acknowledged.

[2] Notice that the perspective distortion is most visible in the first images (wide angle) and diminishes towards the end of the sequence (longer focal length).

## References

[1] P. Beardsley, P. Torr and A. Zisserman, 3D Model Acquisition from Extended Image Sequences, *Proc ECCV'96*, vol.2, pp.683-695.

[2] L. Falkenhagen: Depth Estimation from Stereoscopic Image Pairs Assuming Piecewise Continuous Surfaces. European Workshop on Combined real and synthesic image processing for broadcast and video productions, Nov. 94, Hamburg. Springer Verlag, Germany.

[3] R. Koch, Automatische Oberflächenmodellierung starrer dreidimensionaler Objekte aus stereoskopischen Rundum-Ansichten, PhD thesis, University of Hannover, 1996.

[4] Y. Chen and G. Medioni: Object Modeling by Registration of Multiple Range Images. Proc. Int. Conf. on Robotics and Automation, Sacramento CA, pp. 2724-2729, 1991.

[5] M. Pollefeys, L. Van Gool, M. Proesmans, Euclidean 3D reconstruction from image sequences with variable focal lengths, *Proc. ECCV'96*

[6] M. Pollefeys, L. Van Gool, A. Oosterlinck. The modulus constraint: a new constraint for self-calibration, *Proc. ICPR'96.*

[7] M. Pollefeys and L. Van Gool, A Stratified Approach to Metric Self-Calibration, *Proc. CVPR'97.*

[8] M. Pollefeys and L. Van Gool, Self-Calibration from the Absolute Conic on the Plane at Inifinity, *Proc. CAIP'97.*

[9] P.H.S. Torr, Motion Segmentation and Outlier Detection, *Ph.D. Thesis*, Oxford 1995.