Robust Linear Auto-Calibration of a Moving Camera from Image Sequences

Thorsten Thormählen, Hellward Broszio, and Patrick Mikulastik

University of Hannover, Information Technology Laboratory,
Schneiderberg 32,
30167 Hannover, Germany
{thormae, broszio, mikulast}@tnt.uni-hannover.de
http://www.digilab.uni-hannover.de

Abstract. A robust linear method for auto-calibration of a moving camera from image sequences is presented. Known techniques for auto-calibration have problems with critical motion sequences or biased estimates. The proposed approach uses known linear equations that are weighted by variable factors. Experiments show, that this modification reduces problems with critical motion sequences and that the estimates are not biased. Therefore, the proposed approach is more robust and achieves a higher estimation accuracy.

1 Introduction

Estimation of camera motion and structure of rigid objects using camera images from multiple views is a common task in computer vision and of interest for many applications. This paper considers the case where the camera performs a translational as well as rotational motion.

For the estimation of the camera motion the real camera is represented by a parametric model, which describes the mapping of the observed three-dimensional rigid objects in the two-dimensional image plane of the camera. The parameters of the camera model can be divided into internal and external camera parameters. External camera parameters describe the position and orientation of the camera in space. Internal camera parameters describe aspects of mapping, e.g. the focal length or the position of the principal point. If the internal camera parameters are known, the camera is calibrated. If the camera is not calibrated, it can be described by the projective camera model. The parameters of the projective camera model are combinations of internal and external camera parameters [4, 5, 2].

In order to estimate the parameters of the projective camera model most approaches establish corresponding feature points in the images. During the estimation of the camera parameters, 3D object points are estimated simultaneously. The resulting reconstruction of projective camera views and object points is determined only up to a global projective transformation. This is sufficient for some applications, for example the synthesis of new views [1]. However, in most applications, the projective reconstruction must be transferred into a metric reconstruction. Therefore, the unknown global projective transformation is reduced to an unknown global metric transformation, which corresponds to a determination of the internal camera parameters and the plane at infinity. Their automatic determination from the parameters of the projective camera is called *auto-calibration*.

Early publications assumed that the internal camera parameters are constant over the image sequence. In 1992 Maybank and Faugeras [12, 3] used the equations of Kruppa [10]. The method was developed further [8, 22, 11]. In 1997 Triggs [21] presented the *Absolute Dual Quadric* (ADQ), which was later used by Pollefeys et al. [17, 14, 15] for auto-calibration with variable internal camera parameters. An alternative approach first determines the plane at infinity and afterwards the internal camera parameters. The search range for the plane at infinity in the projective space can be limited by the fact that all observed object points must be located in front of the camera [6, 7, 16, 13].

Our approach is a modification of the linear ADQ approach by Pollefeys et al. In [15] Pollefeys et al. weight the linear equations by the reciprocal of the assumed standard deviations of the internal camera parameters. This incorporation of a priori knowledge reduces the problem with critical motion sequences [20, 9, 19]. However, constraining all internal parameters with fixed weights causes biased estimates, even if it is not necessary, e.g. in cases of sequences without critical motion.

In this paper we try to overcome this disadvantage by introducing linear estimation with variable weights instead of fixed weights.

The following Section briefly reviews Pollefeys' approach with fixed weights. In Section 3 the proposed approach with variable weights is presented. Chapter 4 compares results of the different approaches and conclusions are drawn in Section 5.

2 Linear Auto-Calibration using the Absolute Dual Quadric

Starting point of the auto-calibration algorithm is a projective reconstruction with $k = 1 \ldots K$ projective camera views given by the 3×4 camera matrices $\mathtt{A}_k$ and $j = 1 \ldots J$ object points given by the 4-vectors $\mathbf{P}_j$ in homogeneous coordinates.

Auto-calibration determines the projective 4×4 matrix $\mathtt{T}$, that transforms the projective camera $\mathtt{A}_k$ into a metric camera $\mathtt{A}_k^M$:

$$\mathtt{A}_k^M = \mathtt{A}_k \, \mathtt{T} \qquad \forall \quad k \tag{1}$$

and the object points $\mathbf{P}_j$ of the projective reconstruction into metric object points $\mathbf{P}_j^M$:

$$\mathbf{P}_j^M = \mathtt{T} \, \mathbf{P}_j \qquad \forall \quad j \quad . \tag{2}$$

Whereby a metric camera matrix can be factorized as follows:

$$\mathtt{A}^M = \mathtt{K}\,\mathtt{R}\,[\,\mathtt{I}\,|\, -\mathbf{C}\,] \quad . \tag{3}$$

The 3×3 rotation matrix $\mathtt{R}$ represents the orientation and the 3-vector $\mathbf{C}$ represents the position of the camera. $\mathtt{K}$ is the calibration matrix with

$$\mathtt{K} = \begin{bmatrix} f & s & c_x \\ 0 & r\,f & c_y \\ 0 & 0 & 1 \end{bmatrix} \quad , \tag{4}$$

where f is the focal length, $(c_x, c_y)^\top$ is the principal point offset from the image center, r is the aspect ratio of pixels and s is the skew parameter. The skew s of a real camera is known to be zero. Furthermore, we assume, that the aspect ratio r is known.

In order to determine $\mathtt{T}$, the ADQ $\mathtt{Q}^*_\infty$ is estimated by solving the following auto-calibration equation for all camera views k:

$$\mathtt{A}_k \; \mathtt{Q}^*_\infty \; \mathtt{A}_k^\top \sim \mathtt{K}_k \; \mathtt{K}_k^\top = \omega_k^* \qquad \forall \quad k \quad , \tag{5}$$

where $\mathtt{Q}^*_\infty$ is a 4×4 matrix with rank 3. The 3×3 matrix ω_k^* represents the dual image of the absolute conic (see [5] for details).

In the first step of the linear estimation algorithm the camera matrices are normalized

$$\mathtt{A}'_k = \mathtt{K}_N^{-1} \mathtt{A}_k \tag{6}$$

with

$$\mathtt{K}_N = \mathrm{diag}\left[N_x + N_y, \;\; \frac{1}{r}(N_x + Ny), \;\; 1 \right] \quad , \tag{7}$$

where N_x is the width and N_y is the height of the camera image. Consequently, the normalized auto-calibration equation is

$$\mathtt{A}'_k \; \mathtt{Q}^*_\infty \; \mathtt{A}_k'^\top \sim \mathtt{K}_N^{-1} \mathtt{K}_k \; \mathtt{K}_k^\top \mathtt{K}_N^{-\top} = \omega_k'^* \qquad \forall \quad k \; . \tag{8}$$

After the normalization step the focal length of the normalized camera is $f' \approx 1$ and the principal point offset $(c'_x, \, c'_y)^\top \approx (0, \, 0)^\top$. Pollefeys assumes the standard deviations of the unknown normalized parameters to

$$f' \approx 1 \pm 3 \tag{9}$$

$$c'_x \approx 0 \pm 0.1 \tag{10}$$

$$c'_y \approx 0 \pm 0.1 \quad . \tag{11}$$

From Eq. (8) follows:

$$\omega_k'^* = \begin{bmatrix} f'^2 + c_x'^2 & c'_x \, c'_y & c'_x \\ c'_x \, c'_y & f'^2 + c_x'^2 & c'_y \\ c'_x & c'_y & 1 \end{bmatrix} \approx \begin{bmatrix} 1 \pm 9.01 & \pm 0.01 & \pm 0.1 \\ \pm 0.01 & 1 \pm 9.01 & \pm 0.1 \\ \pm 0.1 & \pm 0.1 & 1 \end{bmatrix} \quad . \tag{12}$$

The symmetrical 4×4 matrix of the ADQ can be parameterized with 10 elements:

$$\mathtt{Q}^*_\infty = \begin{bmatrix} q_1 & q_2 & q_3 & q_4 \\ q_2 & q_5 & q_6 & q_7 \\ q_3 & q_6 & q_8 & q_9 \\ q_4 & q_7 & q_9 & q_{10} \end{bmatrix} \quad . \tag{13}$$

In order to estimate the elements of $\mathtt{Q}^*$, for each camera view 6 linear equations from the following 6 conditions can be derived. Each linear equation is weighted according to its assumed standard deviations from Eq. (12):

$$\omega'^{*}_{12} = 0 \;\Rightarrow\; \tfrac{1}{0.01}\,(\mathbf{a}'_1\ \mathtt{Q}^*_\infty\ \mathbf{a}'^\top_2) = 0 \tag{14}$$
$$\omega'^{*}_{13} = 0 \;\Rightarrow\; \tfrac{1}{0.1}\,(\mathbf{a}'_1\ \mathtt{Q}^*_\infty\ \mathbf{a}'^\top_3) = 0 \tag{15}$$
$$\omega'^{*}_{23} = 0 \;\Rightarrow\; \tfrac{1}{0.1}\,(\mathbf{a}'_2\ \mathtt{Q}^*_\infty\ \mathbf{a}'^\top_3) = 0 \tag{16}$$
$$\omega'^{*}_{11} = \omega'^{*}_{22} \;\Rightarrow\; \tfrac{1}{0.2}\,(\mathbf{a}'_1\ \mathtt{Q}^*_\infty\ \mathbf{a}'^\top_1 - \mathbf{a}'_2\ \mathtt{Q}^*_\infty\ \mathbf{a}'^\top_2) = 0 \tag{17}$$
$$\omega'^{*}_{11} = \omega'^{*}_{33} \;\Rightarrow\; \tfrac{1}{9.01}\,(\mathbf{a}'_1\ \mathtt{Q}^*_\infty\ \mathbf{a}'^\top_1 - \mathbf{a}'_3\ \mathtt{Q}^*_\infty\ \mathbf{a}'^\top_3) = 0 \tag{18}$$
$$\omega'^{*}_{22} = \omega'^{*}_{33} \;\Rightarrow\; \tfrac{1}{9.01}\,(\mathbf{a}'_2\ \mathtt{Q}^*_\infty\ \mathbf{a}'^\top_2 - \mathbf{a}'_3\ \mathtt{Q}^*_\infty\ \mathbf{a}'^\top_3) = 0\,, \tag{19}$$

where $\mathbf{a}'_1, \mathbf{a}'_2, \mathbf{a}'_3$ are the rows of the normalized camera matrix $\mathtt{A}'$.

If the number of camera views is at least 3, an over-determined linear set of equations for the elements of $\mathtt{Q}^*_\infty$ can be generated, which is solved by singular value decomposition [18]. The searched transformation $\mathtt{T}$ can be determined by a singular value decomposition of $\mathtt{Q}^*_\infty$:

$$\begin{aligned} \mathtt{Q}^*_\infty &= \mathtt{U}\ \mathrm{diag}[w_1,\, w_2,\, w_3,\, w_4]\ \mathtt{V}^\top \\ \mathtt{T} &= [\mathtt{U}_3\ \mathrm{diag}[\sqrt{w_1},\, \sqrt{w_2},\, \sqrt{w_3}] \mid (0,0,0,1)^\top]\,, \end{aligned} \tag{20}$$

where the columns of the 4×3 matrix $\mathtt{U}_3$ are those three columns of the 4×4 matrix $\mathtt{U}$, which do not correspond to the smallest singular value w_4.

3 Linear Auto-Calibration with Variable Weights

In order to improve the above algorithm, we propose to use variable weights for Eqs. (18) and (19) instead of the fixed values:

$$\text{Eq. (18)} \;\Rightarrow\; \tfrac{1}{\beta}\,(\mathbf{a}'_1\ \mathtt{Q}^*_\infty\ \mathbf{a}'^\top_1 - \mathbf{a}'_3\ \mathtt{Q}^*_\infty\ \mathbf{a}'^\top_3) = 0 \tag{21}$$
$$\text{Eq. (19)} \;\Rightarrow\; \tfrac{1}{\beta}\,(\mathbf{a}'_2\ \mathtt{Q}^*_\infty\ \mathbf{a}'^\top_2 - \mathbf{a}'_3\ \mathtt{Q}^*_\infty\ \mathbf{a}'^\top_3) = 0 \tag{22}$$

with

$$\beta = 0.1\, e^{(0.3\, n)} \quad . \tag{23}$$

The modified linear algorithm is executed $N = 50$ times with $n = 0$ to $(N-1)$.

By altering β exponentially, it is possible to cover a wide range of weights. If $n = 0 \;\Rightarrow\; \beta = 0.1$, and therefore Eqs. (21) and (22) are considered approximately as much as Eqs. (15)-(17) in the linear equation set. If $n = 49 \;\Rightarrow\; \beta = 242174.76$, and the influence of Eqs. (21) and (22) is negligible.

Changing the weight of Eqs. (18) and (19) correspond to changing the assumed standard deviation of the normalized focal length f' in Eq. (9). Another possibility would be to alter the weights of Eqs. (14) to (17), which would correspond to a change of the assumed standard deviation of the principal point offset in Eqs. (9) and (10). However, this would yield the same results, because the result of the equation set is not changed by a global scale and therefore only the ratio of the assumed standard deviations is important.

Since the modified linear algorithm is executed 50 times with different weights, there are 50 possible solutions for T. Each solution is evaluated by the non-linear cost function, which is proposed by Nistér [13]:

$$\phi = \sum_k \frac{s(\mathtt{A}_k\mathtt{T})^2 + c_x(\mathtt{A}_k\mathtt{T})^2 + c_y(\mathtt{A}_k\mathtt{T})^2 + (r(\mathtt{A}_k\mathtt{T}) - r)^2}{f(\mathtt{A}_k\,\mathtt{T})^2} \tag{24}$$

where the functions $s(.)$, $c_x(.)$, $c_y(.)$, $r(.)$ and $f(.)$ extract respectively the parameters skew, principal point offset in x- and y-direction, pixel aspect ratio and focal length from the camera matrix by QR-decomposition [18]. Finally, the solution with the smallest cost ϕ is selected.

4 Results

4.1 Synthetic Data Experiments

In this subsection two experiments with synthetically generated input data are presented. The first experiment simulates a critical camera motion, that is close to a degenerated case, and the second experiment simulates a non-critical camera motion.

For each experiment 500 synthetic test sequences with random scenes are generated. The random scenes consist of 6000 3D object points, which have a distance from the camera between 36 and 72 mm. Each test sequence consists of 10 images. Approximately 160 to 170 of the object points are visible in each camera image. The errors in the positions of the generated 2D image feature points obey an isotropic Gaussian distribution with standard deviation σ. The camera image has 720×576 pixel and a physical size of 7.68×5.76 mm, thus the pixel aspect ratio is 1.06667. The focal length is 10.74 mm. Principle point offset and skew of the camera are zero. All intrinsic camera parameters are kept constant over the sequence.

In experiment 1 translation and rotation between two successive views are very small (see Tab. 1).

	Translation [mm]	Rotation [deg]
Exp. 1	X = 0.25	pan = -0.05
	Y = 0.0	tilt = -0.075
	Z = 0.05	roll = 0.005
Exp. 2	X = 2.0	pan = -2.0
	Y = 0.0	tilt = -0.5
	Z = 1.0	roll = 0.05

Table 1. Camera motion between two successive views for experiment 1 and 2

Fig. 1 shows the results of experiment 1 for five different standard deviations σ of the position errors of generated 2D feature points. The mean and the standard deviation of the estimation results for all intrinsic camera parameters are plotted. Three different approaches for linear auto-calibration using the ADQ are compared: (#1) The approach with fixed weights described in Sec. 2, (#2) the classical approach that does not weight its linear equations and builds its equation set only with Eqs. 14-17, and (#3) the proposed approach with variable weights.

From Fig. 1a the disadvantage of the approach (#1) with fixed weights is evident. The estimation results for the focal length are pulled to the assumed value of

$$\begin{aligned} N_x + N_y &= (720 + 576)\,\text{pixel} \\ &= (7.68 + 5.76)\,\text{mm} = 13.44\,\text{mm} \end{aligned} \tag{25}$$

by Eqs. (18) and (19). Consequentially, the estimation is biased.

On the other hand, if σ is high, approach (#1) gives much better results for all intrinsic parameters (Fig. 1a-e) than approach (#2). The higher robustness against critical camera motions of approach (#1) is due to the additional equations 18 and 19, which are not used by approach (#2). The proposed approach (#3) with variable weights always performs best.

In experiment 2 translation and rotation between two successive views is large and not close to a critical camera motion (see Tab. 1). Thus, the classical approach (#2) gives good estimation results (Fig. 2). Therefore, the biased estimation results of approach (#1) are unnecessary in this case. In contrast, the estimation results of the proposed approach (#3) with variable weights are as good as the results of approach (#2).

4.2 Natural Image Sequences

The proposed linear auto-calibration approach has also demonstrated to work well on natural image sequences taken by a moving camera. Results of augmented image sequences that have been calibrated using the technique described in this paper are illustrated in Fig. 3. Videos of these augmented image sequences and executables of our non-commercial camera tracker can be found on our website[1].

5 Conclusion

As shown by the experiments the proposed linear auto-calibration approach has nearly no estimation bias and reduces the problem with critical motion sequences. Therefore, it is more robust and achieves an overall higher estimation accuracy than existing approaches.

A slight disadvantage of the proposed approach is its approximately $N = 50$ times higher computational effort. In practice however, this causes no problem, because the computational effort of the linear auto-calibration is small compared to the effort for feature tracking, outlier elimination and estimation of a projective reconstruction. Nevertheless, in future work, it can be tried to reduce N, e.g. by a more explicit detection of critical camera motions.

[1] http://www.digilab.uni-hannover.de

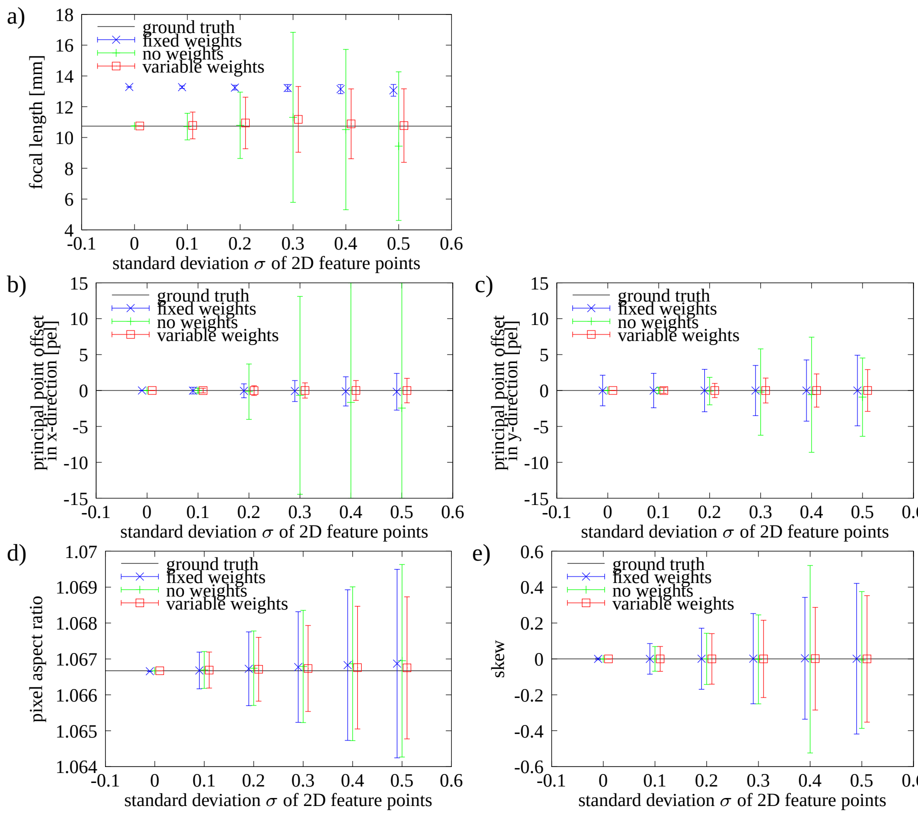

Fig. 1. Results of experiment 1 (critical camera motion): Fig. a)-e) show the ground truth and estimation results of the different approaches for all 5 intrinsic camera parameters over 5 different standard deviations of the position errors of generated 2D feature points. The small symbols mark the mean and the errorbars indicate the standard deviation of the estimation results over 500 random trials.

References

1. Chen, Q.: Multi-view Image-Based Rendering and Modeling. Dissertation, University of Southern California (2000)
2. Faugeras, O., Luong, Q.T.: The Geometry of Multiple Images : The Laws That Govern the Formation of Multiple Images of a Scene and Some of Their Applications. MIT Press (2001)
3. Faugeras, O., Luong, Q.T., Maybank, S.J.: Camera self-calibration: Theory and experiments. In: ECCV. Volume 558 of Lecture Notes in Computer Science. (1992) 321–334
4. Faugeras, O.: Three-Dimensional Computer Vision. MIT Press (1993)
5. Hartley, R.I., Zisserman, A.: Multiple View Geometry. Cambridge University Press (2000)
6. Hartley, R.I.: Cheirality invariants. In: DARPA Image Understanding Workshop. (1993) 745–753
7. Hartley, R., Hayman, E., Agapito, L., Reid, I.: Camera calibration and the search for infinity. In: ICCV. (1999) 510–517

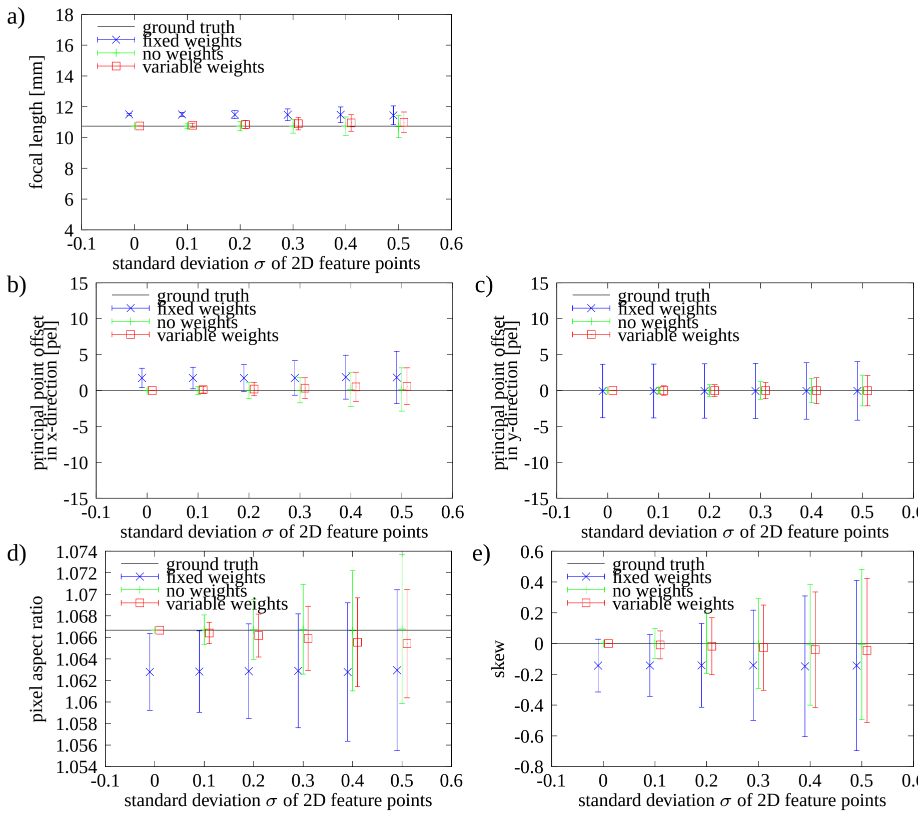

Fig. 2. Results of experiment 2 (non-critical camera motion): Fig. a)-e) show the ground truth and estimation results of the different approaches for all 5 intrinsic camera parameters over 5 different standard deviations of the position errors of generated 2D feature points. The small symbols mark the mean and the errorbars indicate the standard deviation of the estimation results over 500 random trials.

8. Heyden, A., Åström, K.: Euclidean reconstruction from constant intrinsic parameters. In: International Conference on Pattern Recognition. Volume 1. (1996) 339–343
9. Kahl, F., Triggs, B.: Critical motions in euclidean structure from motion. In: CVPR. Volume 2. (1999) 367–372
10. Kruppa, E.: Zur Ermittlung eines Objektes aus zwei Perspektiven mit innerer Orientierung. Sitz-Ber. Akad. Wiss., Wien, Math. Naturw. Abt. IIa. **122** (1913) 1939–1948
11. Luong, Q.T., Faugeras, O.D.: Self-calibration of a moving camera from point correspondences and fundamental matrices. International Journal of Computer Vision **22** (1997) 261–289
12. Maybank, S., Faugeras, O.: A theory of self-calibration of a moving camera. International Journal of Computer Vision **8** (1992) 123–151
13. Nistér, D.: Calibration with robust use of cheirality by quasi-affine reconstruction of the set of camera projection centres. In: ICCV. Volume 2. (2001) 116–123
14. Pollefeys, M., Gool, L.V., Vergauwen, M., Cornelis, K., Verbiest, F., Tops, J.: Video-to-3d. In: Proceedings of Photogrammetric Computer Vision 2002 (ISPRS Commission III

Symposium), International Archive of Photogrammetry and Remote Sensing. Volume 34. (2002) 252–258
15. Pollefeys, M., Gool, L.V., Vergauwen, M., Verbiest, F., Cornelis, K., Tops, J., Koch, R.: Visual modeling with a hand-held camera. International Journal of Computer Vision **59** (2004) 207–232
16. Pollefeys, M., Koch, R., Gool, L.V.: A stratified approach to metric self-calibration. In: CVPR. (1997) 407–412
17. Pollefeys, M., Koch, R., Gool, L.V.: Self-calibration and metric reconstruction in spite of varying and unknown internal camera parameters. In: ICCV. (1998) 90–95
18. Press, W.H., Flannery, B.P., Teukolsky, S.A., Vetterling, W.T.: Numerical Recipes in C, 2nd ed. Cambridge Univ. Press (1992)
19. Sturm, P.: A case against kruppa's equations for camera self-calibration. IEEE Transactions on Pattern Analysis and Machine Intelligence **22** (2000) 1199–1204
20. Sturm, P.: Critical motion sequences for monocular self-calibration and uncalibrated euclidean reconstruction. In: CVPR. (1997) 1100–1105
21. Triggs, B.: Autocalibration and the absolute quadric. In: CVPR. (1997) 609–614
22. Zeller, C.: Calibration projective, affine et euclidienne en vision par ordinateur et application a la perception tridimensionnelle. Dissertation, École Polytechnique (1996)

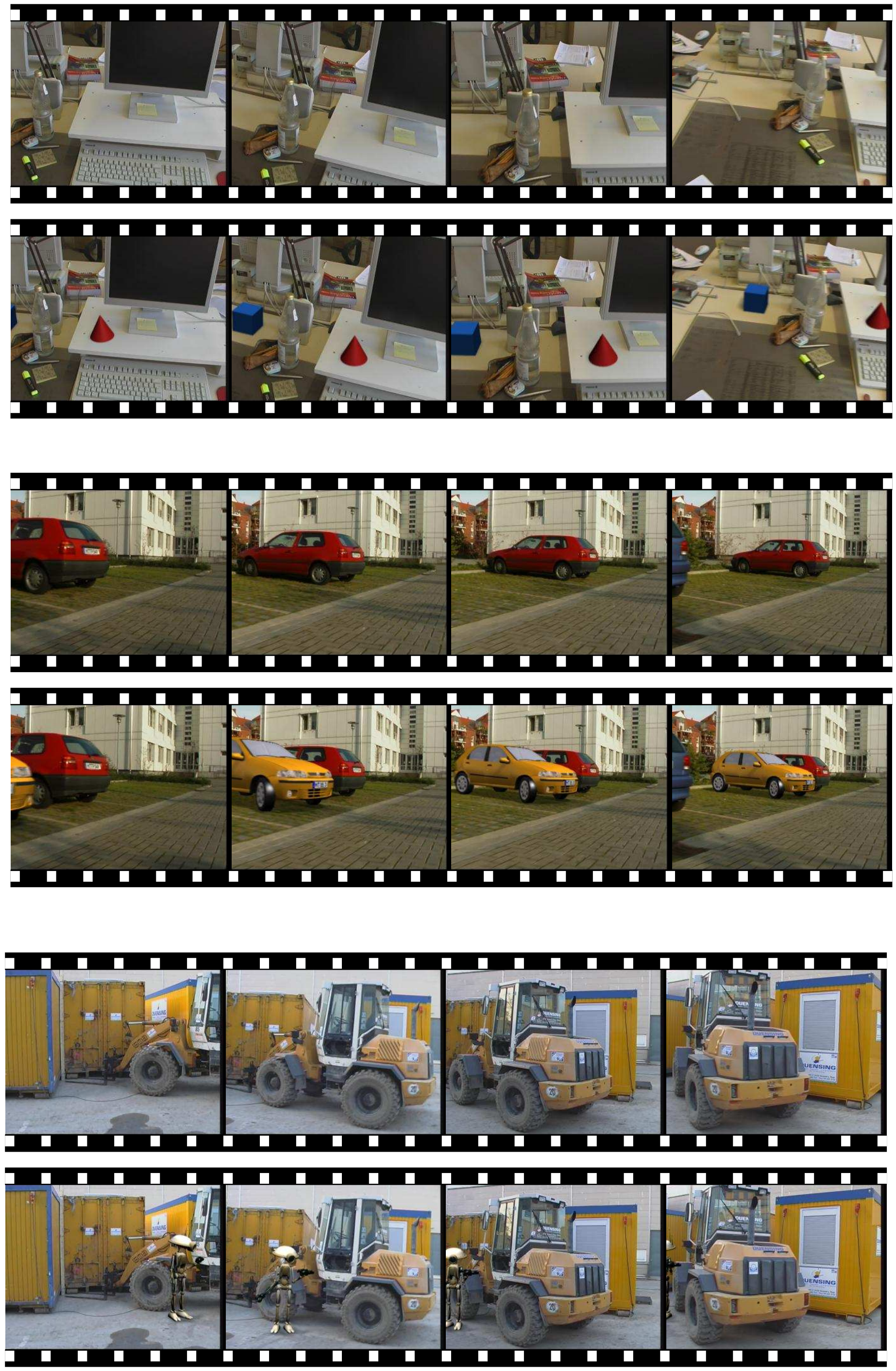

Fig. 3. Examples of augmented image sequences.