

LINEAR TWO VIEW MULTIPLE PLANE GEOMETRY

M. Kirchhof¹

¹ FGAN-FOM, Gutleuthausstr. 1, 76275 Ettlingen, Germany, kirchhof@fom.fgan.de

In this contribution we present a method for direct linear multiple plane estimation and optimization. The application is the detection of independent motion by background subtraction or temporal differences using image stabilization. Given a set of consistent homographies and a corresponding image segmentation for an image pair it is possible to improve the image stabilization by locally warping. We show that linear constraints can be induced to homography estimation to ensure consistency. In this context consistency means that the homographies correspond to the same relative orientation but to different 3d planes. The relative orientation can be derived by measurements of additional sensors e.g. inertial navigation system, odometry or can be computed from the fundamental or essential matrix. The method is compared to standard estimation of homographies with simulated data and the capability of the method is shown using depth estimation on the Middlebury-Stereo Benchmark dataset.

Introduction

Since motion in an image sequence is usually utilized as one of the important cues for security surveillance, object detection and tracking and motion analysis, one can consider motion detection and segmentation as a basic problem in computer vision. Although it has been studied for several decades and various considerable methods have been proposed, it is still of growing interest. In general motion analysis has to be very robust when applied to unconstrained environment. Many approaches rely on optical flow computation [3, 8]. Background subtraction [6, 7] and temporal differences [4] are appropriate methods for motion detection. Applications like vehicle-borne or airborne video surveillance, object detection and tracking based on a moving platform requires initial motion compensation [4, 6, 7].

The optimal motion compensation can be computed from a 3d reconstruction and re-projection of every image point from the textured 3d model at the cost of high computational effort. A more efficient method is to approximate the image motion induced by the sensors motion by an affine mapping or a homography. This is sufficient

for short base lines between the views. But increasing the base line increases the model violation as well until it can no longer be neglected. This effect can be compensated by multiple-homography estimation for each pair of frames. The corresponding 3d planes can become very small resulting in inaccuracy of the homographies. Zelnik-Manor and Irani [9] overcome this problem by rank constraints to a multi-frame-multi-plane-collineation-matrix.

In contrast to this we introduce linear consistency constraints between pairs of homographies. In this context consistency means that the homographies are induced by the same relative orientation between the two views.

In the next section we introduce a method for direct linear computation of one or more homographies enforcing consistency with the relative orientation given by a fundamental matrix. We will show that this procedure is more robust against outliers than standard homography estimation. At the end of this section we prove that this method gives a least square estimate for the algebraic error assuming an exact given relative orientation. We proceed with the introduction of topological constraints which are enforced iteratively with non-linear optimization. Finally we present some experiments with synthetic and real data and discuss the results.

Notation

The image points are generally represented in homogeneous coordinates. The projection of the i -th point in the j -th frame is written as

$${}^j x_i = \begin{bmatrix} {}^j u_i & {}^j v_i & {}^j w_i \end{bmatrix}^T. \quad (1)$$

Estimates for the image points (for example by a given homography) are written as $\tilde{{}^j x}_i$. The cross-product of two arbitrary three-vectors $x_1 = [u_1 \ v_1 \ w_1]^T$ and $x_2 = [u_2 \ v_2 \ w_2]^T$ is written as $x_1 \times x_2 = S_{x_1} x_2 = -x_1^T S_{x_2}$ where S_{x_i} is the corresponding skew symmetric matrix. As long as we do not distinguish between different homographies for a given pair of frames ($j, j+1$) the homographies are notated as ${}^j H$. When the homographies have to be distinguished we use the indexing ${}^j H_k$ for the k -th homography.

Multiple homography estimation

In this section we introduce a method for direct linear computation of multiple homographies consistent to each other. Two homographies are defined to be consistent to each other, if they rely on the same relative orientation but on different 3d planes. A set of homographies is defined to be consistent if each pair of homographies in this set is consistent. It is easy to prove that this condition ensures that all homographies rely on a unique relative orientation.

Therefore the procedure of estimating multiple consistent homographies is equivalent to estimate homographies consistent to a given relative orientation. Please note that it would not be sufficient that one homography is consistent to all the others because the homography decomposition is a quadratic problem and therefore two solutions exist.

Linear constrains to homographies

A linear least square solution requires a linear formulation of the constraints. The homography ${}^j H$ characterizes a mapping

between image points in the j -th to the $j+1$ -th frame by

$${}^{j+1} \tilde{x}_i \equiv {}^j H {}^j x_i, \quad (2)$$

where “ \equiv ” denotes equal up to scale. Equation (2) is equivalent to ${}^{j+1} x_i \times {}^j H {}^j x_i \equiv 0_{3 \times 1}$ which gives us two linear and independent equations for $H^j = [{}^j h_1 \ {}^j h_2 \ {}^j h_3]$ per correspondence.

Therefore four correspondences determine the homography ${}^j H$ up to scale. This is sufficient because homographies are homogeneous [2].

A fundamental or essential matrix gives the bilinear constraint

$${}^{j+1} x_i^T {}^j F {}^j x_i = 0. \quad (3)$$

It can also be computed from the relative orientation measured by additional sensors by

$${}^j F = K^T \begin{bmatrix} {}^j R S_{({}^j R^{-1}({}^j t))} \end{bmatrix} K, \quad (4)$$

where ${}^j R$ represents the relative rotation from the camera position in frame j to frame $j+1$ and ${}^j t$ the relative translation measured in the coordinate system attached to the camera position in frame j and K is the calibration matrix of the camera [2].

Combination of equations (2) and (4) yields to

$${}^j x_i^T {}^j H^T {}^j F {}^j x_i = 0. \quad (5)$$

While this holds for every ${}^j x_i$ the matrix ${}^j H^T {}^j F$ is skew symmetric. This constraint can be expressed by the linear expression

$${}^j H^T {}^j F + {}^j F^T {}^j H = 0_{3 \times 3} \quad (6)$$

giving us five linear independent homogeneous equations for ${}^j H$. One might think that two additional correspondences would be sufficient to determine a unique homography. But this is not correct. Each correspondence gives only one independent equation in addition to the constraint (6). This is because the constraint (6) already enforces that each correspondence $({}^{j+1} x_i, {}^j x_i)$ fulfills the epipolar constraint: ${}^{j+1} x_i$ is positioned on the line $F {}^j x_i$.

Implementation

The estimation of the consistent homographies is implemented by an adaptive RANSAC (RANdom SAMple Consensus) algorithm combined with guided matching. Hypothesis

for homographies consistent with a given relative orientation (computed from the images or measured by additional sensors) represented by the fundamental matrix are computed from equation (6) and three correspondences using equation (2). After finding the hypothesis with the largest support we compute a least square solution from all inliers to the current hypothesis.

The least square solution can not be obtained from Equation (6) and Equation (2) for all inliers; because the constrained (6) would be treated as soft-constrained.

Therefore we compute the set of solutions $\langle {}^jH_1^*, \dots, {}^jH_4^* \rangle$ to Equation (6) which is a 4-dimensional subspace in the domain of homographies. The set of solutions can be obtained from singular value decomposition of Equation (6). Then we compute the least square solution to (2) in the 4-dimensional subspace: Let ${}^jh_1^*, \dots, {}^jh_4^*$ be a basis of the 4-dimensional subspace by composing the columns of the solutions ${}^jH_1^*, \dots, {}^jH_4^*$ to a vector. Let further jA be the action-matrix designed from (2) by using the complete inlier set. Then the least square solution to (2) holding the constrained (6) is given by the solution to

$$\underset{|\alpha|=1}{\operatorname{argmin}}({}^jA[{}^jh_1^*, \dots, {}^jh_4^*]\alpha), \quad (7)$$

which can be computed with singular value decomposition. The final result is then given by the back-substitution

$${}^jh = [{}^jh_1^*, \dots, {}^jh_4^*]\alpha. \quad (8)$$

Topology and non-linear optimization

The next step is the labeling of the image that is used for piecewise warping. First the correspondences are labeled with the index of the homography that minimizes the re-projection error. The pixel-wise labeling is then based on nearest neighbor calculation: Each pixel is labeled with the index of the closest correspondence.

A key problem of this procedure is that no topological constraints can be used. Therefore we keep only the labeling of the

largest topologic connected area. For a set of k different homographies we introduce the k -dimensional state-vector ${}^jS_k(i)$ normalized by its infinity-norm. The l -th component of ${}^jS_k(i)$ represents the membership of the correspondence $({}^{j+1}x_i, {}^jx_i)$ to the l -th homography. If the correspondence $({}^{j+1}x_i, {}^jx_i)$ is labeled with l the l -th component of ${}^jS_k(i)$ is set to one. Correspondences $({}^{j+1}x_i, {}^jx_i)$ that are not labeled get a uniform state vector. Finally all state-vectors, homographies and the relative orientation are optimized by minimizing

$${}^j\varepsilon = \sum_i {}^j\varepsilon_i^T \operatorname{diag}({}^jS_k(i)) {}^j\varepsilon_i, \quad (9)$$

using Levenberg-Marquardt optimization [2]. Here the l -th component of ε_i equals the residual projection error for the homography jH_l : $\|{}^{j+1}\widetilde{x}_{li} - {}^{j+1}x_i\|$ with ${}^{j+1}\widetilde{x}_{li} \equiv {}^jH_l {}^jx_i$. The homogeneous points ${}^{j+1}\widetilde{x}_{li}$ and ${}^{j+1}x_i$ are normalized by their third component. $\operatorname{diag}(\cdot)$ converts a vector to a diagonal matrix. Note that during the optimization the state vectors are treated as continuous variables.

Experiments

The robustness of the algorithm is compared to standard homography estimation with a synthetic scene. The scene consists of about 100000 3d points located in three planes and a volume that is not occluded by the planes. 20 different baselines were tested. Figure 1 shows that the proposed algorithm requires about half of the RANSAC runs than standard homography estimation.

Figure 2 shows the distribution of the true inliers of the estimation procedures and its standard deviations. As mentioned above the proposed algorithm captures much more inliers than standard homography estimation. Only for completeness we included Figure 3 which shows the orientation error of the normal vector in degree. Since the proposed algorithm has to estimate only the normal vector the residuals are much smaller.

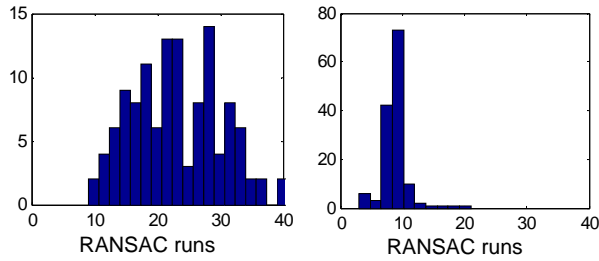


Fig.1 Histogram of required RANSAC runs. Left: standard homography estimation; Right: Proposed method with constraints on the homography.

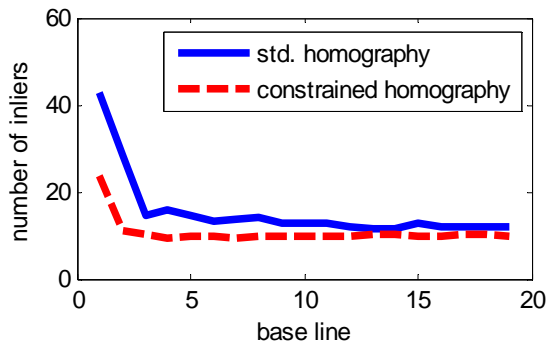
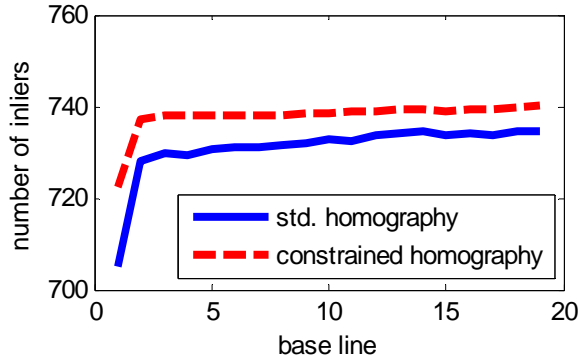


Fig.2 Distribution of true inliers. Top: mean value; Bottom: standard deviation.

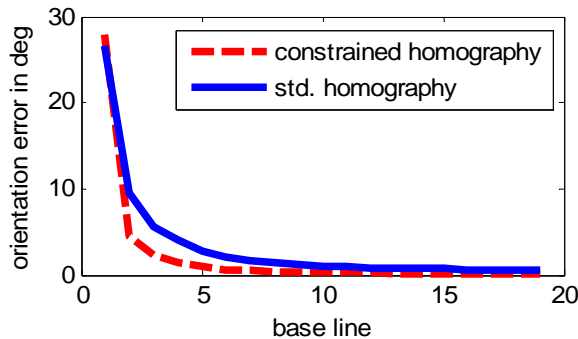


Fig.3 Mean of the orientation error of the normal vector in degree.

The algorithm was assessed with the Middlebury-Stereo Benchmark dataset. Figure 4 shows the results of the evaluation: The third row shows the result of a classical dense stereo technique using dynamic programming [1] while the fourth row shows

the results of the proposed method. Large disparity is displayed bright while black indicates that the disparity could not be computed. The disparity map of the proposed algorithm is computed from the state-vector combined with the estimated disparity for the corresponding correspondence. The result of the second dataset shows a large gap at the lower right corner. This occurs because the object lies in the same 3d plane as a larger object in the middle of the frame.

The algorithm has correctly rejected the smaller area. The effect can be solved by computing additional local homographies for such areas. The result was not improved to show how the algorithm works.

Conclusion

We have presented a novel method for direct linear estimation of consistent sets of homographies. The consistency was enforced by linear constraints using the fundamental matrix. A valid topology was ensured by the selection of the largest labeled regions combined with non-linear optimization of the back projection error based on the established correspondences.

Since the algorithm is currently implemented in MATLAB, a meaningful number for the computational time could not be given. Nevertheless it is known that structure from motion, which is of the same computational effort, is possible in real time.

Similar to the work of Woelke and Koch [8] the results can be improved by computing the state-vector pixel-wise in boundary-regions. This should result in smooth boundaries between different labeled regions.

The topological constraints can although be enforced by Markov-Random-Fields. This statistical modeling could improve the result but at the cost of high computational effort.

References

1. L. Falkenhagen, 1997. *Hierarchical Block-Based Disparity Estimation Considering Neighbourhood Constraints*. Proceedings International Workshop on SNHC and 3D Imaging, 115-122.

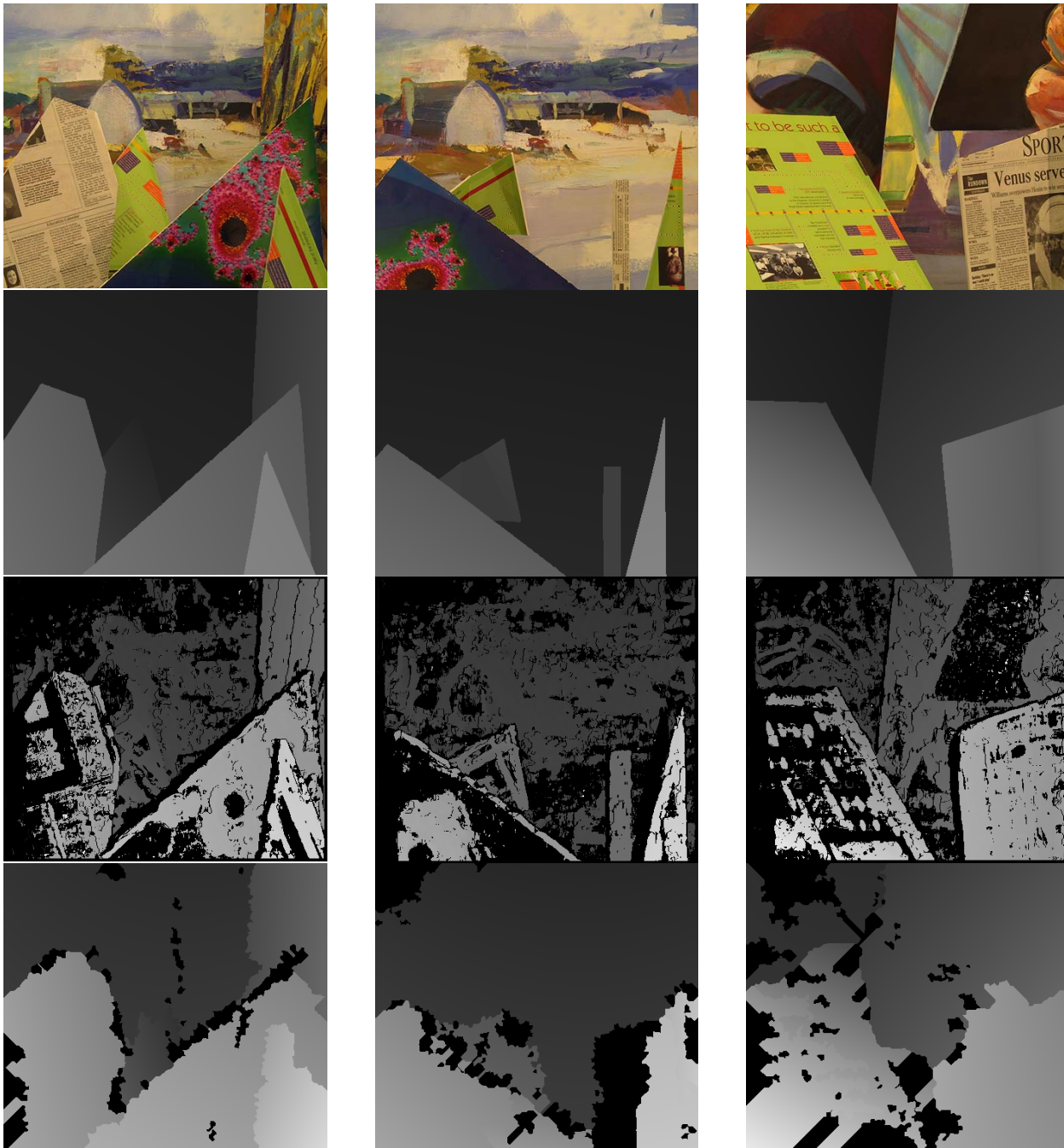


Fig. 4. From top to down: one frame of the dataset, ground truth disparity map, disparity map computed with dynamic programming, and the outcome of the proposed algorithm.

2. R. Hartley, A. Zisserman, 2000. *Multiple View Geometry*. Cambridge University Press, Cambridge, UK.
3. J. Kang, I. Cohen, G. Medioni, C. Yuan, 2005. *Detection and Tracking of Moving Objects from a Moving Platform in Presence of Strong Parallax*. In Proceedings of the IEEE International Conference on Computer Vision, 1, 10 – 17.
4. M. Kirchhof, U. Stilla, 2006. *Detection of moving objects in airborne thermal videos*. ISPRS Journal of Photogrammetry and Remote Sensing, Volume 61, Issues (3-4), 187-197.
5. Y. Ren, C.-S. Chua, Y.-K. Ho, 2003. *Statistical background modelling for non stationary camera*. Pattern Recognition Letters 24, pp. 183-196
6. C. Stauffer, W.E.L. Grimson, 1999. *Adaptive background mixture models for real-time tracking*. CVPR, Vol. 2, 246-252
7. F. Woelk, R. Koch, 2004. *Fast monocular Bayesian detection of independently moving objects by a moving observer*. In: Proceedings of DAGM Symposium. Lecture Notes in Computer Science, Vol. 3175, 27–35.
8. L. Zelnik-Manor, M. Irani, 2002. *Multiview Constraints on Homographies*. IEEE Transaction on Pattern Recognition and Machine Intelligence 24(1)