

LINEAR CONSTRAINTS IN TWO-VIEW MULTIPLE HOMOGRAPHY ESTIMATION OF UNCALIBRATED SCENES

Michael Kirchhof

FGAN-FOM Research Institute for Optronics and Pattern Recognition
Gutleuthausstrasse 1, 76275 Ettlingen, Germany
-kirchhof@fom.fgan.de

Commission III, WG III/1

KEY WORDS: Geometry, Vision, Homography, Parameter estimation, Video

ABSTRACT:

In this contribution, we present a method for direct linear estimation of multiple homographies and optimization of estimation results by enforcing topological constraints by means of connected components. The applications are the detection of independent moving objects seen by a moving observer and sparse reconstruction of the scene. Security and surveillance applications require day-and-night capability. Therefore thermal cameras are the preferred sensors. These sensors violate primary constraints for the estimation of the optical flow. The motion of the sensor is compensated by image stabilization, which allows us to utilize results from the extensive field of change detection, for example, background subtraction and temporal differences. We assume that the scene can be approximated by a composition of several planes. Each plane induces an independent homography for each image pair. A set of homographies for one image pair is called consistent if and only if all homographies refer to the same relative orientation but even though they may refer to different 3d-planes. We show that linear constraints can be induced to homography estimation to ensure consistency within a pair of views. The fundamental matrix allows us to formulate a solution independent of the calibration of the camera. Given a set of consistent homographies and a corresponding image segmentation for an image pair it is possible to improve the image stabilization by local warping instead of warping with one single homography. This approach is explored by experiments on disparity map estimation based on the homography estimation on the Middlebury-Stereo Benchmark dataset. Additionally, the robustness of the algorithm is explored using a synthetic scene. Finally we show first results of motion detection based on this method combined with temporal differences.

1. INTRODUCTION

Motivation. Homography estimation is used for 3D analysis, mosaicing, camera calibration and more. Our focus is the stabilization of image sequences for the estimation of independent motion. 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 methods have been proposed, it is still of considerable interest. The typical sensors for visual security surveillance are thermal cameras. Because of their low local contrast and geometric resolution they require special algorithms different from optical flow (Kirchhof and Stilla, 2006). On the other hand, many vision applications deal with a small depth relief of the observed scene compared with the extent of the image. Often these scenes can be approximated by a plane, e.g., the ground plane. A homography-based approach is appropriate for modeling such configurations. For urban environments, however, this is not sufficient. Manmade environments consist essentially of planes. Therefore, we assume that the scene can be described by a composition of several planes. This leads to the estimation of multiple homographies for a single pair of views.

The estimation of multiple homographies for one pair of views can be implemented by successive estimation of single homographies followed by the exclusion of the inlier for the

next estimation step. The main problem is that the independent estimation of multiple homographies is very instable due to small supporting regions. The most common solution to this is to introduce nonlinear constraints that represent the rigidness of the motion and the scene (Zelnik-Manor and Irani, 2002, Kaehler and Denzler, 2007). But these robustifications rely on an initial guess of the homographies followed by nonlinear optimization. In contrast, we propose a set of linear constraints to the homography estimation that represent the rigid motion only. Therefore, we are able to propose a linear estimation procedure for multiple homographies that rely on the same rigid motion.

Related Work. In general, motion analysis has to be very robust when applied to unrestricted environments. Many approaches rely on optical flow computation (Kang et al., 2005, Woelk and Koch, 2004). Background subtraction (Ren et al., 2003, Stauffer and Grimson, 1999) and temporal differences (Kirchhof and Stilla, 2006) are appropriate methods for motion detection. Applications like vehicle-borne or airborne video surveillance and object detection and tracking based on a moving platform require initial motion compensation (Ren et al., 2003, Kirchhof and Stilla, 2006, Stauffer and Grimson, 1999). Optimal motion compensation can be computed from a 3d reconstruction and reprojection 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 baselines between the

views. But increasing the baseline increases the model violation to the point where 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, 2002). The estimation can be made more robust by the use of additional knowledge, for example, the rigidness of the motion and the scene.

The conventional way to introduce the rigidness constraints is to parametrize the homography H_{lk} by geometrically relevant parameters. The mapping of the image point \hat{x}_i in the l -th frame to the k -th frame can be expressed by

$$\lambda_i \hat{x}_i^k = \mathbf{K} \left(\mathbf{R}_{kl} - t_{kl} n_i^\top / d_i \right) \mathbf{K}^{-1} \hat{x}_i^l = H_{lk} \hat{x}_i^l. \quad (1)$$

The index kl denotes a mapping with respect to a relative orientation from the k -th to the l -th frame. The inner orientation of the camera is represented by \mathbf{K} . The relative orientation is given by the rotation matrix \mathbf{R}_{kl} , the translation t_{kl} is represented in the coordinate system of the k -th frame and the 3d plane is represented by the normalized normal vector n_i and the distance to the origin d_i observed in the coordinate system of the l -th frame. λ_i is a nonzero scale factor. The rotation \mathbf{R}_{kl} and the translation t_{kl} constitute the relative orientation between the two cameras with five parameters since only the ratio t_{kl}/d_i is determinable from equation (1). The decomposition of H_{lk} according to equation (1) is possible only if the calibration matrix \mathbf{K} is known and has up to eight solutions. These eight solutions can be reduced to two reasonable solutions (Faugeras and Lustman, 1988) by ensuring that the 3d points lie in front of both cameras. Now it is possible to parametrize the homographies between two views by a global relative orientation (six parameters for translation and rotation) valid for every homography and the corresponding 3d planes (three parameters per plane), which leads to a minimal parametrization of the problem. (Note that in contrast to a single coplanarity assumption one has to deal with six instead of five parameters for the relative orientation). This minimal geometric parametrization can be used for multiple frames and one single 3d plane (Kirchhof and Stilla, 2006) as well as multiple planes visible in multiple views (Kaehler and Denzler, 2007).

A different way to treat multiple planes observed in multiple frames is to introduce rank constraints on a factorized composition of homographies to a multi collineation matrix (Zelnik-Manor and Irani, 2002, Shashua and Avidan, 1996). This method requires computation of a scaling factor for each correspondence and each homography since the homography gives only homogeneous relations between correspondences (equation 1). Contribution We introduce a method for direct linear computation of one or more homographies enforcing consistency with the relative orientation given by a fundamental matrix without knowledge of the calibration. We will show that this procedure is more robust against outliers than standard homography estimation. The method gives a least squares estimate for the algebraic error assuming an exact given relative orientation. We proceed with a nonlinear optimization of the result with enforcement of the topological constraint of connected components. In practice the relative orientation is computed from correspondences as well and is therefore also

optimized during the non linear optimization. During the optimization the assignment of the correspondences to the multiple homographies is improved and the reprojection error is minimized.

Implementation overview While estimating consistent homographies direct linear is at the focus of this contribution we like to give only a brief overview of the whole estimation procedure. We start with the detection of points of interest (POI) with the Förstner-operator (Förstner and Gülich, 1987). The correspondences between subsequent frames are then established with the KLT-tracker (Shi and Tomasi, 1994). At this stage the correspondences are erroneous. Therefore a robust filtering technique based on an adaptive random sample consensus (RANSAC) (Fischler and Bolles, 1981, Hartley and Zisserman, 2000) is used to enforce the epipolar constraint presented in equation (2). Depending on the availability of calibration data the fundamental matrix respectively the essential matrix is estimated using the linear 8 point algorithm (Hartley and Zisserman, 2000) or the 5 point algorithm (Nistér, 2004). From the surviving correspondences we can now estimate consistent homographies for a single pair of frames by enforcing the epipolar constraint (2) on the homographies as presented in section 2 using RANSAC again. The iterative procedure of estimating homographies is stopped if to few correspondences are left or if the last homography had very low support. The resulting putative correspondences for each plane are an initial guess for the nonlinear optimization presented in section 2 which also takes care of the connected components constraints. The established plane based segmentation of the image can now be used for various applications such as dense depth estimation Figure 7 or motion detection Figure 6.

Notation For formulation and representation, we use the framework of algebraic projective geometry. Homogeneous vectors and matrices will be denoted by upright boldface letters, e.g., \mathbf{x} or \mathbf{H} and Euclidean vectors or matrices with slanted boldface letters, e.g., \mathbf{x} or \mathbf{H} . In homogeneous coordinates, '=' means an assignment or an equivalence up to a scaling factor $\lambda \neq 0$. Some parameters have to be represented in various coordinate systems. Observations in the coordinate system S_k attached to the k -th frame are denoted by an overhead index, e.g., $\hat{\mathbf{x}}_i^k$. Relative orientations representing the motion from S_k to S_l or mappings between the two frames k and l are written as $(\mathbf{R}_{kl}; t_{kl})$. We also have to represent various homographies for a pair of frames $k; l$. We use \mathbf{H}_{kl} for the i -th homography consistent with the fundamental matrix F and \mathbf{H}_{kl}^{pi} for the homography that relies on the i -th 3D plane π_i^k .

2. METHODOLOGY

In the first paragraph of this section, we introduce a method for direct linear computation of multiple homographies consistent with each other. Two homographies that rely on different 3d planes are defined to be consistent to each other, if they rely on the same relative orientation. 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 estimating homographies consistent with a given

relative orientation. Note that it would not be sufficient that one homography is consistent to all the others because homography decomposition is a quadratic problem and therefore two reasonable solutions exist (Faugeras and Lustman, 1988). In the second paragraph, we introduce our optimization procedure. During the optimization, the topological constraint of connected components is enforced. This constraint is motivated by the idea that we have to deal with a rigid scene that can be composed of several planar surface elements. The constraint is designed to prevent that any virtual 3D plane is fitted through the whole scene without any physical complement.

Constraints for multiple homography estimation In this paragraph, we show that linear constraints can be induced to homography estimation to ensure consistency within a pair of views. The linear constraints for the homographies between two views can be obtained from consistency with the corresponding fundamental matrix \mathbf{F}_{lk} , which only states the relative orientation

$$0 = \overset{l}{\mathbf{x}}_i^T \mathbf{K}^{-T} \mathbf{R}_{lk} [t_{lk}]_x \mathbf{K}^{-1} \overset{l}{\mathbf{x}}_i = \overset{l}{\mathbf{x}}_i^T \mathbf{F}_{lk} \overset{l}{\mathbf{x}}_i. \quad (2)$$

The second part of the equation is well known as epipolar constraint. As stated before, more details on fundamental matrices and their estimation can be found in (Hartley and Zisserman, 2000). Combining the equations (1 and 2) results in the consistency constraint

$$0 = (\mathbf{H}_{lk})^T \mathbf{F}_{lk} \overset{l}{\mathbf{x}}_i. \quad (3)$$

Equation (3) is valid for every point $\overset{l}{\mathbf{x}}_i$ of frame l even if the correspondence $(\overset{l}{\mathbf{x}}_i, \overset{k}{\mathbf{x}}_i)$ is not an inlier of the homography \mathbf{H}_{lk} . This implies that the matrix $\mathbf{H}_{lk}^T \mathbf{F}_{lk}$ is skew symmetric, which can be expressed by the matrix equation

$$0 = \mathbf{H}_{lk}^T \mathbf{F}_{lk} + \mathbf{F}_{lk}^T \mathbf{H}_{lk}. \quad (4)$$

The advantage of the representation, that we present here, is that the constraints of equation (3) are homogeneous and linear. While equation (4) is symmetric, it gives us five linear independent homogeneous equations, that can be solved using singular value decomposition. The solution is a set of four homographies

$$\mathcal{SH}_{lk} = \{\overset{1F}{\mathbf{H}}_{lk}, \dots, \overset{4F}{\mathbf{H}}_{lk}\}, \quad (5)$$

which are consistent with the fundamental matrix \mathbf{F}_{lk} . Each homography that is consistent with the fundamental matrix can be represented by a linear combination of \mathcal{SH}_{lk}

$$\mathbf{H}_{lk} = \sum_{i=1}^4 \overset{iF}{\mathbf{H}}_{lk} \alpha_i. \quad (6)$$

This linear combination can be estimated in a direct linear transformation model from the tracked image features with random sample consensus (RANSAC) (Fischler and Bolles, 1981, Hartley and Zisserman, 2000). First we compute the action matrix from a set of correspondences $(\overset{l}{\mathbf{x}}_i, \overset{k}{\mathbf{x}}_i)$ the same way as for standard homography estimation by

$$0 = (\overset{k}{\mathbf{x}}_i \times \mathbf{H}_{lk} \overset{l}{\mathbf{x}}_i) \iff 0 = \mathbf{A}_H \mathbf{h}_{lk}, \quad (7)$$

where \mathbf{h}_{lk} is composed from the columns of the homography \mathbf{H}_{lk} . Note that equation (7) is equivalent to equation (1). Then the direct linear estimation of the homography without constraints can be computed by solving equation (7) using singular value decomposition (Hartley and Zisserman, 2000). This procedure was used as comparison to our approach in the experiments in section 3.

The linear constraints can now be enforced by multiplying the action matrix with composition of the columns of the set of homographies \mathcal{SH}_{lk}

$$0 = \mathbf{A}_H [\overset{1F}{\mathbf{h}}_{lk} \overset{2F}{\mathbf{h}}_{lk} \overset{3F}{\mathbf{h}}_{lk} \overset{4F}{\mathbf{h}}_{lk}] \alpha \quad (8)$$

and solving for the multi-index using singular value decomposition. The resulting homography can now be computed from the linear combination (6). The resulting homography is a least square solution of (7) with respect to constraint (4). A proof can be found in Appendix 5 of ((Hartley and Zisserman, 2000)).

Alternatively the constraint can be embedded the same way as the constraints in (Shashua and Avidan, 1996, Zelnik-Manor and Irani, 2002, Kaehler and Denzler, 2007, Kirchhof and Stilla, 2006). This means that the constraints are included in a nonlinear optimization after the initial homography estimation. Since the constraints are almost linear, the procedure will converge to the global minimum in one step.

Topology and nonlinear optimization The fundamental matrix gives us a set of correspondences, hopefully no longer disturbed by outliers. The estimation of multiple homographies gave us an initial set of homographies and a initial partition of the set of correspondences by assigning each correspondence to the homography that fits best. In this step we update this partition in terms of connected components. In the same step we can optimize all homographies and the relative orientation with respect to (4).

The first step toward the topological constraint of connected components is the labelling of the image. First the correspondences are labelled with the index of the homography that minimizes the re-projection error. Note that at this point correspondences are used only if they are an inlier for at least one homography. The pixel-wise labelling is then based on a nearest-neighbor calculation: Each pixel is labelled with the index of the closest correspondence. The key problem of this

procedure is that topological constraints can not be enforced direct. Therefore we keep only the labelling of the largest connected area. For each correspondence $(\overset{\downarrow}{\mathbf{x}}_i, \overset{\downarrow}{\mathbf{x}}_i)$ we introduce the $t+1$ dimensional state vector $\mathbf{S}_{lk}^t(i)$, where t is the number of homographies found for the two images. The j -th component of $\mathbf{S}_{lk}^t(i)$ represents the membership of the $(\overset{\downarrow}{\mathbf{x}}_i, \overset{\downarrow}{\mathbf{x}}_i)$ to the j -th homography. If the correspondence $(\overset{\downarrow}{\mathbf{x}}_i, \overset{\downarrow}{\mathbf{x}}_i)$ is labelled with j , the j -th component of $\mathbf{S}_{lk}^t(i)$ is set to 1. Correspondences that are not labelled receive a 1 in the $t+1$ -th component of the state vector $\mathbf{S}_{lk}^t(i)$. Finally, all state-vectors, homographies and the relative orientation are optimized by minimizing

$$\epsilon = \sum_i (\epsilon_{lk}(i)^T \text{diag}(\mathbf{S}_{lk}^t(i)) \epsilon_{lk}(i) + d(\overset{\downarrow}{\mathbf{x}}_i, \mathbf{F}_{lk} \overset{\downarrow}{\mathbf{x}}_i)^2) \quad (9)$$

with respect to equation (4) using Levenberg-Marquardt optimization (Hartley and Zisserman, 2000). Here the j -th component of $\epsilon_{lk}(i)$ equals the reprojection error for the homography $\overset{\pi_j}{\mathbf{H}}_{lk}$:

$$\epsilon_{lk}(i) = \left\| \overset{\downarrow}{\mathbf{x}}_i - \overset{\downarrow}{\mathbf{x}}_i \right\| \quad (10)$$

with $\overset{\downarrow}{\mathbf{x}}_i = \overset{\pi_j}{\mathbf{H}}_{lk} \overset{\downarrow}{\mathbf{x}}_i$. The homogeneous points $\overset{\downarrow}{\mathbf{x}}_i$ and $\overset{\downarrow}{\mathbf{x}}_i$ are normalized by their third component. $\text{diag}(\cdot)$ converts a vector to a diagonal matrix. $d(\mathbf{a}, \mathbf{b})$ gives the geometric distance between the point \mathbf{a} and line \mathbf{b} . Note that during the optimization the state vectors are treated as continuous variables. Therefore the state-vector $\mathbf{S}_{lk}^t(i)$ has to be normalized by its infinity-norm during the optimization to match the interpretation of a probability distribution.

3. EXPERIMENTS

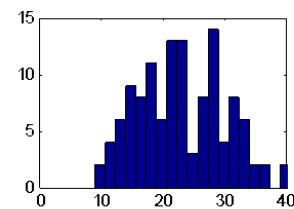
The robustness of the algorithm is compared to standard homography estimation with a synthetic scene. The scene consists of about 100000 3d points located in three planes and a volume that is not occluded by the planes. 20 different baselines were tested. Figure 1 shows that the proposed algorithm requires about 50% fewer RANSAC runs than standard homography estimation with the same parameter setting for the adaptive RANSAC. The adaptive RANSAC estimates the number N of required runs by

$$N = \frac{\ln(1-p)}{\ln(1 - (\text{inlier complexity}))} \quad (11)$$

where *inlier* denotes the fraction of inlier and *complexity* the complexity of the model (4 for standard homography estimation and 3 for constraint homography estimation). Note that in this

case complexity is the minimal size of a set of correspondences that define a unique hypothesis and not the number of estimated parameters. This choice of N ensures that with probability p ($p = 0.99$ in all experiments) at least one of the random samples was free from outliers. More important than the absolute values is the comparison with the theoretical values computed from equation (11) $N \approx 12.1$ for standard homography estimation and $N \approx 8.4$ for the proposed method. Only points located at the largest plane were treated as inliers which gives a percentage of 75%. This indicates that constrained homography estimation is more robust against noise when the convex hull of the hypothesis generating set decreases.

RANSAC runs standard method



RANSAC runs improved method

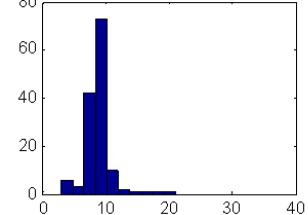


Figure 1: Histogram of required RANSAC runs for estimating one of the three homographies induced by the three planes. Top: standard homography estimation; bottom: Proposed method with constraints on the homography.

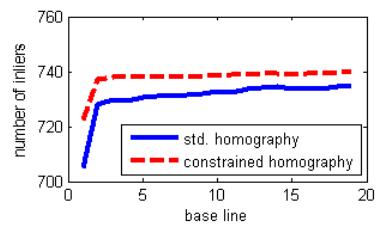


Figure 2: Mean of the distribution of detected true inlier. The exact number was 745.

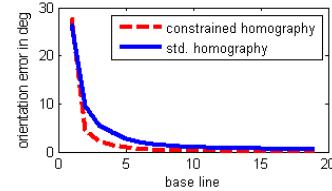


Figure 3: Mean of the orientation error of the normal vector in degrees.

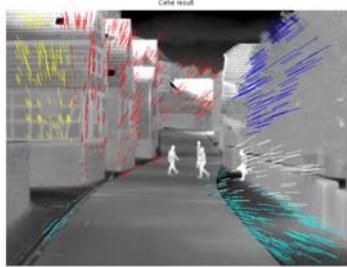


Figure 4: Correspondences between two views are displayed in one of the views. The labelling of the correspondences is colour coded.



Figure 5: Temporal difference between a reference view with a view transformed by a homography.



Figure 6: Temporal difference of a reference view and a view transformed by local warping with multiple homographies.

In the second experiment, the algorithm was assessed with the Middlebury-Stereo Benchmark dataset. Figure 7 shows the results of the evaluation: The third row shows the result of a classical dense stereo technique using dynamic programming (Falkenhagen, 1997) 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 with an estimate of the disparity for the corresponding homography. The result for 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. To show how the algorithm works the result was not improved.

In the final experiment, we demonstrate the capability of the proposed algorithm for independent motion detection. We used a dataset captured with an AIM thermal camera in the $8 - 12 \mu\text{m}$ band. The baseline between the two views was about 5m . Figure 4 shows the labelling of the correspondences with the

best homography. The resulting segmentation was uncertain and some regions were not labelled as well. Therefore we used the minimum of the absolute difference between the transformed and the reference image for every homography for the boundaries between different segments and the unsegmented regions. The outcome of the proposed algorithm in Figure 6 is compared to the stabilization of the scene with one single homography in Figure 5. The bright regions at the buildings at the right side of Figure 5 appear because the estimated homography was fitted to the complete set of correspondences which were very dense at the buildings on the left side of the figure. These bright regions would result in falsely detected motion. The bright regions at the margin of Figure 6 are caused by the forward motion of the sensor. Because of this motion, the transformed frame is much smaller than the reference frame.

4. 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 in terms of connected components of the 3d points corresponding to the same 3d plane projected in the reference frame was ensured by the selection of the largest labelled regions combined with non-linear optimization of the back projection error based on the correspondences established.

Since the algorithm is currently implemented in MATLAB, a meaningful computation time can not be given. Nevertheless it is known that structure from motion, which involves the same computational effort, is possible in real time. At present, the labelling of the pixels is based on nearest neighbours. This often results in frayed boundaries between different segmented regions. Additional image information such as edges should be used to improve the results. In similarity to the approach of (Woelk and Koch, 2004) the results can be improved by computing the state vector pixel wise in boundary regions. This should result in smooth boundaries between different labelled regions. The topological constraints can also be enforced by Markov-Random-Fields. This statistical modelling may improve the accuracy but at the cost of high computational effort.

Obviously the processing is affected by the precision of the fundamental matrix. Without loss of generality we can say that, the precision of the fundamental matrix in terms of the geometric distance between an image point and the corresponding epipolar line has to be higher than the threshold for the homography estimation, because otherwise we would not be able to compute any homography. Since the fundamental matrix is updated during the nonlinear optimization smaller errors in the fundamental matrix would not disturb the overall result as long as the initialization point lies within radius of convergence. A detailed analysis of this radius has to follow.

OUTLOOK

The presented method can not be adapted to multiple frames and one plane. But in a similar way we can build up some additional constraints for multiple planes and multiple frames.

We assume that we have found two homographies $\mathbf{H}^{1,2}$ and $\mathbf{H}^{2,1}$ for one image pair. The line of intersection between the two corresponding 3d planes is mapped identically by both

homographies. Therefore the projection of this line to the l -th

$$\overset{l}{g} = (\overset{1}{\mathbf{H}}_{lk})^{-1} \overset{2}{\mathbf{H}}_{lk} \overset{l}{g} = \overset{1}{\mathbf{H}}_{lk} \overset{2}{\mathbf{H}}_{lk} \overset{l}{g}. \quad (12)$$

Since the 3d line and the projection is the same for the image

frame $\overset{l}{g}$ is invariant under the homography pairs l and $k+1$ the line g is also invariant under the corresponding homography

$$\overset{l}{g} = \overset{1}{\mathbf{H}}_{(k+1)l} \overset{2}{\mathbf{H}}_{l(k+1)} \overset{l}{g}. \quad (13)$$

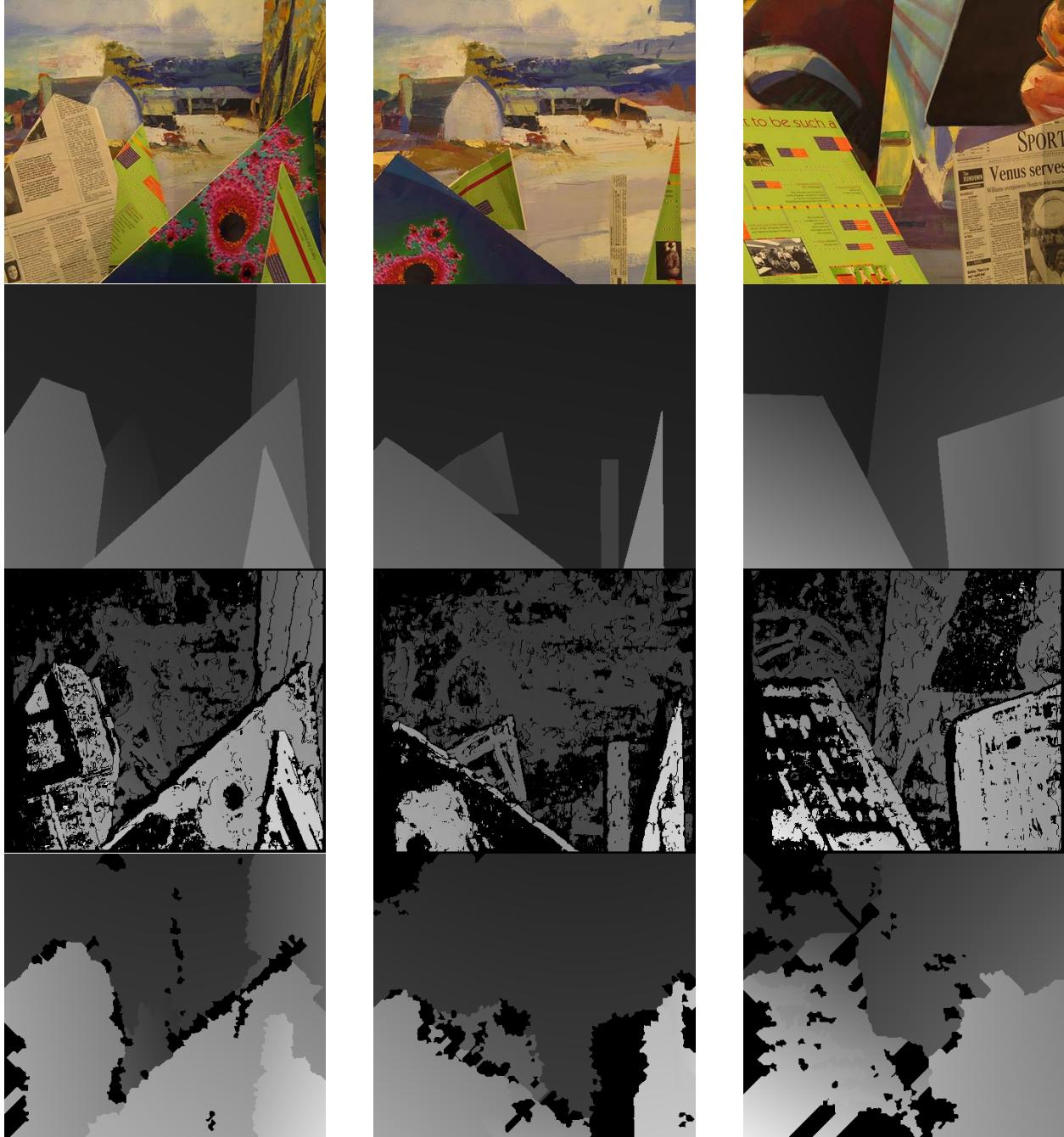


Figure 7: From top to bottom: one frame of the dataset, ground truth disparity map, disparity map computed with dynamic programming and the outcome of the proposed algorithm.

REFERENCES

- Falkenhagen, L., 1997. Hierarchical block-based disparity estimation considering neighborhood constraints. In: *International Workshop on SNHC and 3D Imaging*, pp. 115–122.
- Faugeras, O. and Lustman, F., 1988. Motion and Structure from Motion in a piecewise planar Environment. *International Journal of Pattern Recognition in Artificial Intelligence* 2, pp. 485–508.
- Fischler, M. A. and Bolles, R. C., 1981. Random Sample Consensus: A Paradigm for Model Fitting with Applications to Image Analysis and Automated Cartography. *Communications of the Association for Computing Machinery* 24(6), pp. 381–395.
- Förstner, W. and Gülch, E., 1987. A Fast Operator for Detection and Precise Location of Distinct Points, Corners and Centres of Circular Features. In: *ISPRS Intercommission Workshop*, Interlaken.
- Hartley, R. and Zisserman, A., 2000. *Multiple View Geometry in Computer Vision*. Cambridge University Press, Cambridge.
- Kaehler, O. and Denzler, J., 2007. Rigid motion constraints for tracking planar objects. In: *Proceedings of DAGM, Lecture Notes in Computer Science*, Vol. 4713, pp. 102–111.
- Kang, J., Cohen, I., Medioni, G. and Yuan, C., 2005. Detection and tracking of moving objects from a moving platform in presence of strong parallax. In: *International Conference on Computer Vision*, Vol. 1, pp. 10–17.
- Kirchhof, M. and Stilla, U., 2006. Detection of moving objects in airborne thermal videos. *ISPRS Journal of Photogrammetry and Remote Sensing* 61(3-4), pp. 187–197.
- Nistér, D. 2004. An Efficient Solution to the Five-Point relative Pose Problem. *IEEE Transactions on Pattern Recognition and Machine Intelligence* 26(6), pp. 756–769.
- Ren, Y., Chua, C.-S. and Ho, Y.-K., 2003. Statistical background modeling for non stationary camera. *Pattern Recognition Letters* 24, pp. 183–196.
- Shashua, A. and Avidan, S., 1996. The rank 4 constraint in multiple view geometry. In: *Proceedings of the European Conference on Computer Vision*, pp. 196–206.
- Shi, J. and Tomasi, C., 1994. Good features to track. In: *IEEE Conference on Computer Vision and Pattern Recognition (CVPR'94)*.
- Stauffer, C. and Grimson, W., 1999. Adaptive background mixture models for real-time tracking. In: *International Conference on Computer Vision and Pattern Recognition*, Vol. 2, pp. 246–252.
- Woelk, F. and Koch, R., 2004. Fast monocular bayesian detection of independently moving objects by a moving observer. In: *Proceedings of DAGM, Lecture Notes in Computer Science*, Vol. 3175, pp. 27–35.
- Zelnik-Manor, L. and Irani, M., 2002. Multiview Constraints on Homographies. *IEEE Transactions on Pattern Recognition and Machine Intelligence* 24(2), pp. 214–223.

