DETECTION OF PLANAR PATCHES IN HANDHELD IMAGE SEQUENCES

Olaf Kähler, Joachim Denzler

Friedrich-Schiller-University, Dept. Mathematics and Computer Science, 07743 Jena, Germany
{kaehler,denzler}@informatik.uni-jena.de

KEY WORDS: planar patches, homography, feature detection, degeneracy

ABSTRACT:

Coplanarity of points can be exploited in many ways for 3D reconstruction. Automatic detection of coplanarity is not a simple task however. We present methods to detect physically present 3D planes in scenes imaged with a handheld camera. Such planes induce homographies, which provides a necessary, but not a sufficient criterion to detect them. Especially in handheld image sequences degenerate cases are abundant, where the whole image underlies the same homography. We provide methods to verify, that a homography does carry information about coplanarity and the 3D scene structure. This allows deciding, whether planes can be detected from the images or not. Different methods for both known and unknown intrinsic camera parameters are compared experimentally.

1 INTRODUCTION

The detection and tracking of features is one of the preliminaries for many applications, ranging from motion analysis to 3D reconstruction. Depending on the complexity of features, more or less knowledge can be gained directly from them. The typical approach is to match corresponding point features over an image sequence, which is solved for many applications (Shi and Tomasi, 1994). Inferring information about the 3D structure of the scene can benefit however from additional constraints, e.g. coplanarity of points (Bartoli and Sturm, 2003). In fact planes are relatively easy to handle as features and do have many useful geometric properties.

Planes have caught the interest of research before. Linear subspace constraints on the motion of planes have been elaborated and used for separating independently moving objects (Zelnik-Manor and Irani, 1999). For the representation of video there are many applications related to planes or so called layers, either for efficient coding exploiting the 2D object motion (Baker et al., 1998, Odone et al., 2002), or aimed towards an interpretation of the 3D scene structure (Gorges et al., 2004). The benefits of incorporating coplanarity constraints (Bartoli and Sturm, 2003) or of explicitly using planes for 3D reconstruction (Rother, 2003) have been investigated, too. Also efficient auto-calibration algorithms in planar scenes are possible (Triggs, 1998). More recently many of the above results have been combined to allow explicit tracking of 3D camera motion from image intensities (Cobzas and Sturm, 2005).

Despite many applications, the automatic extraction of planar regions is still a difficult task. The work of Baker (Baker et al., 1998) was one of the first setting the trend to use homographies for finding planes. Later algorithms made use of random sampling to automatically detect points mapped under a common homography (Schindler, 2003). Using a sparse set of tracked point features, random sampling was also applied for a Least Median of Squares regression to detect a *dominant* homography in the scene (Odone et al., 2002). The dominant homography is defined as the one transferring all known points with the least median transfer error. The extraction of dominant homographies is iterated to find smaller and smaller planar patches. A very similar algorithm was given in (Gorges et al., 2004). The dominant homography in that case is defined as the one transferring most points correctly.

The mentioned works concentrate on finding point features or image regions underlying a common homography. This is a necessary condition for the points to reside on the same 3D plane, it is not a sufficient one however. A very simple case is a camera not moving at all between two frames. All points are then transferred with the same homography, the identity matrix. Yet the points may reside in many different 3D scene planes. A similar well known situation occurs, if the camera undergoes a pure rotation. Especially when processing image sequences from handheld or head-mounted cameras, both of these cases are abundant and ignoring them leads to erroneous planes being detected. Detection of planar patches in a scene should therefore not only find image regions under a homography, but also decide, whether coplanarity can be inferred from the detected homographies.

The detection of related degenerate cases is an important issue in many different computer vision tasks, yet rarely addressed in research. A seminal work on the topic (Torr et al., 1999) is considering the case of degeneracy for the estimation of the fundamental matrix. The basic task in that work is to find a guidance for feature matching, either the epipolar geometry or a homography warp on the whole image. This is highly related to our problem and we will develop similar methods in our work.

The rest of the paper is organized as follows. In section 2 we will shortly introduce the notation and present a useful decomposition of homography matrices. Finding homographies from known point correspondences is reviewed in section 3. The task of deciding on coplanarity from given homographies is elaborated in section 4. In section 5 an experimental evaluation of the developed methods is given. Some final remarks on further work and conclusions will sum up the results in the end.

2 PRELIMINARIES

Throughout the work we will use the standard projective camera model projecting world points $\boldsymbol{X}$ onto image points $\boldsymbol{x}$ with $\boldsymbol{x} = \alpha K(R\boldsymbol{X} + \boldsymbol{t})$. The matrix K is an upper triangular matrix containing intrinsic parameters, R is a rotation matrix and $\boldsymbol{t}$ the translation vector. We typically need two camera frames and two sets of camera parameters, which are then denoted with a index, e.g. K_1 and K_2. Restricting to two frames it is sufficient to know the relative motion, and hence we set $R_1 = \mathrm{Id}$, $\boldsymbol{t}_1 = \boldsymbol{0}$ and $R_2 = R$, $\boldsymbol{t}_2 = \boldsymbol{t}$.

A world plane is defined by the inner product $\boldsymbol{n}^T\boldsymbol{X} = d$, with inhomogeneous 3D vectors $\boldsymbol{n}$ and $\boldsymbol{X}$, and a scalar d. Every world plane projected into two images induces a homography between the two images, a 2D-2D projective transformation H. This H maps the projections $\boldsymbol{x}_1$ of world points on the plane onto corresponding points $\boldsymbol{x}_2$ in the second projection. This is easily shown using the relative motions:

$$\begin{aligned}\boldsymbol{x}_2 =& \alpha_2 K_2(R_2\boldsymbol{X} + \boldsymbol{t}_2) = \alpha_2 K_2(R + \frac{1}{d}\boldsymbol{t}\boldsymbol{n}^T)\boldsymbol{X} \\ \boldsymbol{X} =& \frac{1}{\alpha_1} R_1^{-1} K_1^{-1}\boldsymbol{x}_1 - R_1^{-1}\boldsymbol{t}_1 = \frac{1}{\alpha_1} K_1^{-1}\boldsymbol{x}_1 \\ \boldsymbol{x}_2 =& \alpha \underbrace{K_2(R + \frac{1}{d}\boldsymbol{t}\boldsymbol{n}^T)K_1^{-1}}_{=H} \boldsymbol{x}_1 \end{aligned} \qquad (1)$$

The homography matrix H therefore is only defined up to an unknown scale.

3 FROM POINTS TO HOMOGRAPHIES

To detect planar patches we first establish point correspondences between consecutive image frames in the sequence using KLT-tracking (Shi and Tomasi, 1994). As points on planar regions underlie a homography, the first step in finding these planar regions is to establish groups of point features correctly transformed by a common 2D-2D projective transformation. Different approaches to do this mainly use the two concepts of random sampling and iterative dominant homography estimation. Before going into their details in sections 3.2 and 3.3, we will shortly review the computation of homographies.

3.1 Computation of Homographies

In an usual approach the 2D homography can be estimated from 4 point correspondences by solving the following linear equation system for the entries of H:

$$\boldsymbol{x}_2 = \alpha H \boldsymbol{x}_1 \qquad (2)$$

With equality up to scale, each pair of corresponding points leads to 2 independent equations in the entries of H. As the matrix H can only be computed up to scale, it has 8 degrees of freedom. Hence four points determine the entries of the matrix H.

With known epipolar geometry however, even three points are a sufficient minimum parameterization of planes. There are several ways of exploiting this (Hartley and Zisserman, 2003). This will basically enforce the computation of homographies compatible with the epipolar geometry, in the sense that the single globally rigid scene motion stored in the epipolar geometry is enforced to be also valid for all points of the homographies. This will fail however, if there are multiple independent motions. In general also the computation of epipolar geometry frequently gives rise to numerical problems.

For our work we do not use the epipolar constraints, but compute homographies directly from equation 2. We typically use more than 4 points and solve the overdetermined system using SVD techniques.

3.2 Random Sampling

The RANSAC approach was used e.g. in (Schindler, 2003) to detect points underlying the same homography. Basically the idea of model fitting with random sampling is very intuitive. Starting with a minimal set of random samples, which define an instance of the model, the support for this instance among the other available samples is measured. In the end we keep the hypotheses with highest support.

For our homography problem, the algorithm has to randomly select points from all known correspondences, so that the parameters of the homography can be determined. This means three random points with known epipolar geometry or four points in the more general case. The errors for transferring the remaining point correspondences with this homography can be computed. Each point correctly transferred up to e.g. 2 pixels difference can be counted as supporting the hypothesis that the points are coplanar. If only the initially selected points support the hypothesis, these points are most likely not coplanar and the computed homography does not have any physical meaning.

This idea of extending an initial homography H_i to more point correspondences can be applied iteratively. After an extension step, a new homography H_{i+1} can be computed with the additional points included. The new homography matrix H_{i+1} can again be extended to all other points correctly transferred. The iteration ends, if no more points are added to the computations. With this approach the result is more robust against small matching inaccuracies in the initially selected points.

3.3 Iterative Dominant Homography Estimation

In various works (Odone et al., 2002, Gorges et al., 2004) the homography explaining most point correspondences is called the *dominant* homography. To find this dominant homography, first the RANSAC algorithm is applied as above. From all sampled candidates only the single best one is kept however. This is defined to be either the one with Least Median overall transfer error (Odone et al., 2002), or the one transferring the largest number of points correctly (Gorges et al., 2004). This dominant homography of the scene is accepted as a planar region, the covered points are removed and another iteration step is started to find the dominant homography of the remaining points.

For the least median error method, the breakdown point is at 50% outliers. If there are many small planes in the scene each covering only a small portion of the image, the homographies found will thus explain only a small portion of all point correspondences. The homography with least median transfer error is then almost arbitrary, and will not necessarily be exactly valid for any but the initially sampled points used to construct it. We therefore decided not to use the least median error method, but to count the points correctly transferred up to e.g. 2 pixels tolerance instead.

3.4 Locality Constraints

If the mentioned homography detection algorithms are applied as described above, they will mostly detect *virtual* homographies. These are induced by *virtual* planes, i.e. geometrically valid 3D planes with many observable points on them, but without any corresponding physical plane. An example can be seen in Figure 1. Note that from geometry and the computed homographies alone, these virtual planes do well represent sets of coplanar points and there is no way to detect them. Additional constraints to prevent the virtual planes therefore can not result from pure photogrammetry. Two basic approaches occur in the literature.

In the work of (Gorges et al., 2004) an explicit locality criterion is used. Only points in a certain neighborhood region are sampled to compute the initial hypotheses in the RANSAC algorithm. In the extension steps, points outside the boundaries of the local neighborhood can be taken into account as well. This might seem like a

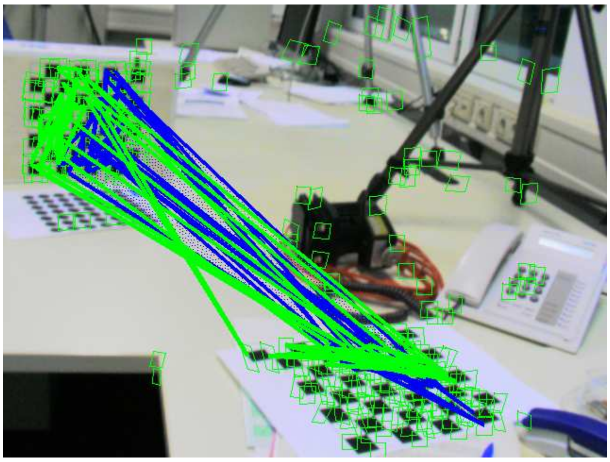

Figure 1: The points connected by the green and blue lines are lying on two virtual planes, which represent coplanar points on planes that do not correspond to any physical object plane

heuristic at first, however it directly facilitates the detection of *locally* planar structures. Starting from the locally planar neighborhood, the iterative extension of the homography to more points still allows the detection of larger planes with arbitrary shape. In our experiments this method practically eliminated the detection of virtual planes.

A more complex but in essence quiet similar criterion was used in (Schindler, 2003). There the plane detection is initialized with equilateral triangles selected by random sampling. All points inside the triangles have to match the same homography, and only then a region growing is started. This is basically an extension of the mentioned locality constraint above, first from an arbitrary shaped neighborhood to the convex interior of a triangle and second from sparse point correspondences to a dense constraint on all image points. Due to the higher complexity with basically the same effect, we have not investigated this method further.

4 FROM HOMOGRAPHIES TO PLANES

Detecting image regions underlying one common homography is only the first step for finding planar patches in an image sequence. All planar image regions will underlie a homography, but not all image regions underlying a homography are necessary coplanar. We will first show that these cases occur exactly if there is no translational motion between the two frames under consideration. Further we will present several methods for detecting these cases in different scenarios, like known or unknown intrinsic camera parameters.

The mentioned problematic cases are directly apparent from the homography decomposition given in equation 1:

$$H = K_2(R + \frac{1}{d}\boldsymbol{t}\boldsymbol{n}^T)K_1^{-1}$$

If the term $\boldsymbol{t}\boldsymbol{n}^T$ vanishes for planes with arbitrary normals $\boldsymbol{n}$, the homographies do not contain any information about the planes, but only consist of $K_2RK_1^{-1}$. On the other hand any homography matrix H containing the second term, has one unique plane with normal $\boldsymbol{n}$ inducing it.

The term vanishes for arbitrary $\boldsymbol{n}$ if and only if $\boldsymbol{t} = 0$. In that case we have a pure rotational motion or change of intrinsic parameters and can not infer anything on the 3D structure. To ensure, a homography does contain relevant information about a 3D plane, we therefore have to test for a translation $\boldsymbol{t} \neq 0$. A first class of testing methods is to analyze a single homography matrix and check it for a particular form. A different class is taking into account additional information from other correspondences.

Algorithms in the first class are testing, whether a given H is of the form $K_2RK_1^{-1}$. Note these methods will always fail to identify the plane at infinity. This is the plane containing all the vanishing points, and it has the normal $\boldsymbol{n} = 0$. So the homography of this plane is always of the form of a pure camera rotation. Only once a translational part is detected in the homography of any other plane, it could be inferred that $\boldsymbol{t} \neq 0$ and hence the homography $H = K_2RK_1^{-1}$ must be induced by the plane with normal $\boldsymbol{n} = 0$.

This inference, like the approaches using knowledge from other correspondences, can only be used in case of a globally rigid motion of the scene however, and not in case of independently moving objects in the scene. This becomes apparent for the example of an object moving in front of a static camera. The plane induced homographies of the object do have a translational motion part, and the whole static background is underlying the same homography. But the background does not necessarily consist of one single plane. If a static scene is assumed on the other hand, the additional information will ease the task of detecting motions without translations.

4.1 Known Intrinsic Parameters

If the intrinsic camera parameters are known, a simple and straight forward test for the translational part in a homography is possible. Multiplying the homography matrix H with the intrinsic parameter matrices K_1 and K_2^{-1} from left and right we get:

$$\begin{aligned} H' = K_2^{-1}HK_1 &= \alpha K_2^{-1}K_2(R + \frac{1}{d}\boldsymbol{t}\boldsymbol{n}^T)K_1^{-1}K_1 \\ &= \alpha(R + \frac{1}{d}\boldsymbol{t}\boldsymbol{n}^T) \end{aligned}$$

It is obvious that the term $\frac{1}{d}\boldsymbol{t}\boldsymbol{n}^T$ vanishes if $\boldsymbol{t} = 0$, i.e. there is no translational part in the camera motion. The larger $\boldsymbol{t}$, the more is H' dominated by a rank-1 part and deviating from the pure rotation matrix R.

A test for H' to be a rotation matrix is given by the singular value decomposition. For the rotation matrix R, all singular values are equal to 1. Taking into account the unknown scale factor α, the ratio of largest to smallest singular value of H' will therefore be 1 if $\boldsymbol{t} = 0$ or $\boldsymbol{n} = 0$. For our experiments we used a slightly less restrictive threshold of 1.2 for the ratio.

4.2 Unknown but Constant Intrinsic Parameters

Needing knowledge of the intrinsic parameters clearly is a shortcoming of the method above. We will consider the next simple case, where the intrinsic camera parameters are unknown, but known to be constant. This scenario is of great practical relevance and has been studied before (Triggs, 1998). Many and especially cheap cameras are not equipped with a zoom-lense and hence fulfill the requirement.

In the case of a constant intrinsic parameter matrix $K = K_1 = K_2$, the homography matrix H is similar (i.e. conjugate) to the matrix $R + \frac{1}{d}\boldsymbol{t}\boldsymbol{n}^T$. This means the two matrices do have the same determinant, eigenvalues and some more properties which are not relevant here, although the singular values might differ.

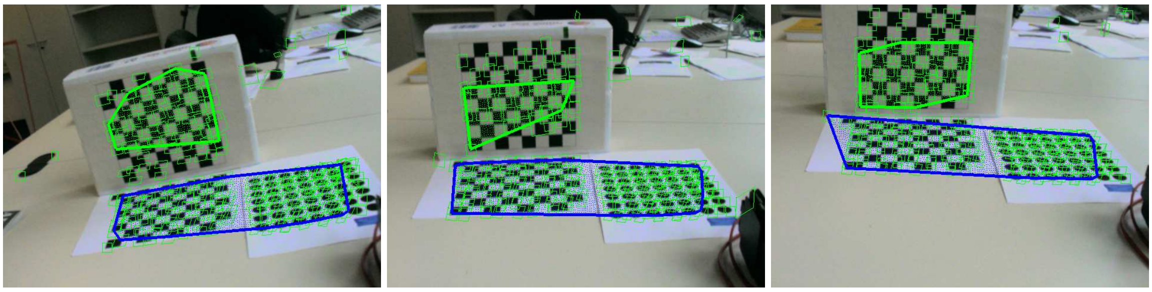

Figure 2: Excerpts of a calibration pattern scene with planar patches detected in the individual frames shown as polygons with thick boundary lines.

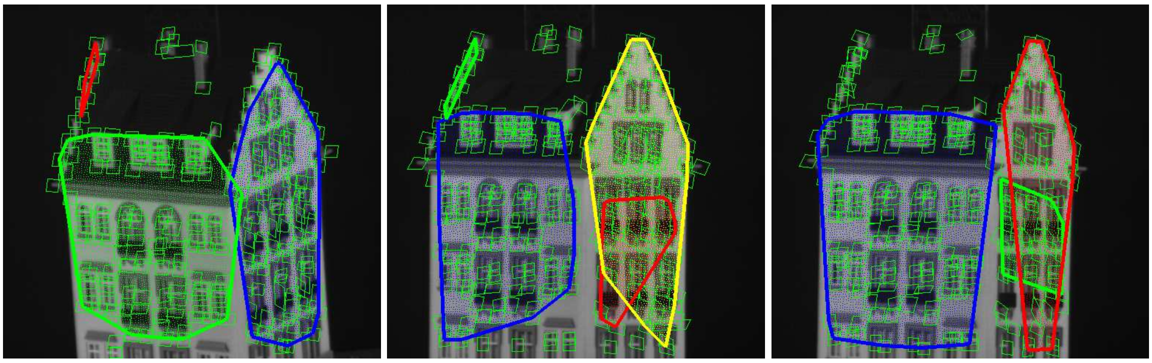

Figure 3: Excerpts of an architectural scene with the thick polygons delineating planar patches found from point correspondences.

The equivalence of eigenvalues is derived from:

$$\begin{aligned}\det(\frac{1}{\alpha}H - \lambda\mathrm{Id}) &= \det(KRK^{-1} + \frac{1}{d}K\boldsymbol{t}\boldsymbol{n}^T K^{-1} - \lambda KK^{-1}) \\ &= \det(K)\det(R + \frac{1}{d}\boldsymbol{t}\boldsymbol{n}^T - \lambda\mathrm{Id})\frac{1}{\det(K)} \\ &= \det(R + \frac{1}{d}\boldsymbol{t}\boldsymbol{n}^T - \lambda\mathrm{Id})\end{aligned}$$

The eigenvalues are given as the roots of this characteristic polynomial and are hence identical for the two matrices. Using this result and the equality $\det(A + \boldsymbol{x}\boldsymbol{y}^T) = (1 + \boldsymbol{y}^T A^{-1}\boldsymbol{x})\det(A)$ it follows, that H has the same eigenvalues up to scale with the rotation matrix R, if and only if $\boldsymbol{n}^T R^T \boldsymbol{t} = 0$. All three eigenvalues of the rotation matrix R do have the same absolute value 1. So do the eigenvalues of the homography matrix H up to the common scale α, if the intrinsic parameters are constant. The ratio of largest to smallest absolute eigenvalue hence provides a means of detecting cases with $\boldsymbol{n}^T R^T \boldsymbol{t} = 0$. In our experiments we again used a ratio of 1.2 as a threshold, to tolerate the effects of slight noise.

The condition tested by this criterion is either met for $\boldsymbol{t} = 0$ or $\boldsymbol{n} = 0$ or if the vectors $R\boldsymbol{n}$ and $\boldsymbol{t}$ are orthogonal. This provides a slightly over-sensitive test for the detection of translations. The case where this measure generates false alarm is a translation in a plane parallel to the plane inducing H.

As mentioned before, these two tests can be extended to a global measure, if we assume a globally rigid motion. Detecting a translational part in any homography matrix, we can assume the whole scene has undergone a translation, and hence every observed homography H carries information about coplanarity. This way the cases where the test is oversensitive can be avoided as well, unless the camera motion is parallel to all planes in the scene.

4.3 Global Homography

Another very intuitive idea exploiting the rigid motion constraint is to simply count, how many points are not correctly transferred between the frames using the homography H. In the case of no translation between the frames, the homography matrix for any plane will be the same. The second, parallax term will vanish and $H = H_\infty = K_2 R K_1^{-1}$. Therefore if all points are transferred with the homography H_∞, the motion of the points was most likely caused by a camera movement without translation. For practical purposes a small portion of outliers should be allowed, depending on the quality of the point correspondences found. In our experiments we considered a homography as global, if it transferred more than 80% of all points with a small transfer error.

However, again there are cases where this test will fail, e.g. if only one plane is visible in the scene. This plane is not necessarily the plane at infinity with $\boldsymbol{n} = 0$, but could as well be a real object plane filling the whole view. Knowledge of the intrinsic parameters and one of the tests above could decide upon this ambiguity.

4.4 Epipolar Geometry

Another way of explicitly using points not residing in the potential plane is to take into account the epipolar geometry. Note with the usual 8-point-algorithm the fundamental matrix F can only be determined up to a two-parameter family of matrices in the case of all points residing in the same 3D plane or no translation occuring between the frames (Torr et al., 1999). Testing for these rank-deficiencies when computing the epipolar geometry will therefore allow the detection of cases without translation.

This test basically has the same restrictions as for the global homography computation before. In fact the same condition that all points underly a common homography is only tested differently here. But again the numerically problematic epipolar geometry is needed, and a small portion of incorrect point correspondences could severely affect this method.

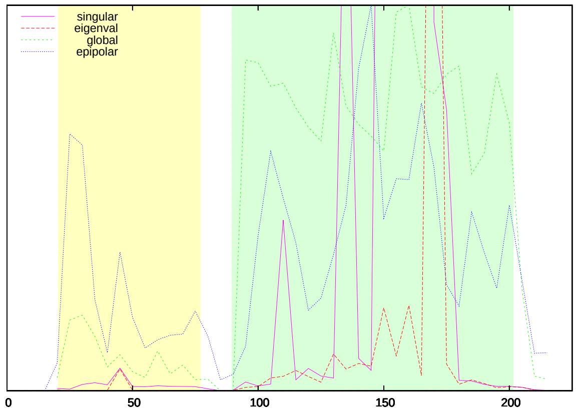

Figure 4: Confidence of different criteria that a translational camera motion was present in the individual frames of a sequence. The yellow background indicates ground-truth frames with pure camera rotation, the green background indicates general motion

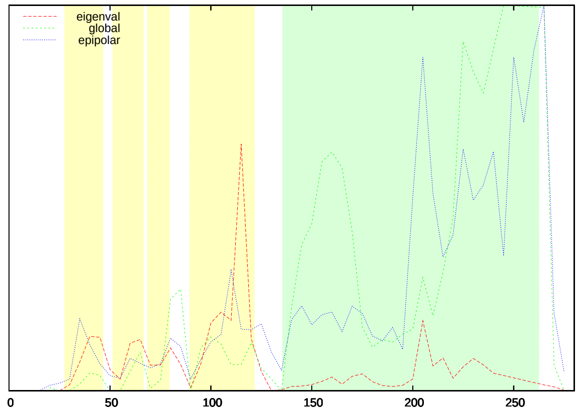

Figure 5: Confidence of different criteria that a translational camera motion was present in the individual frames of a sequence. The yellow background indicates ground-truth frames with purely zooming camera, the green background indicates general motion

5 EXPERIMENTS

We have presented a method for detecting homographies and several different methods for checking the information on planarity contained in a homography. For the experimental evaluation we follow a similar structure. First the results from homography detection are shown qualitatively, as this part of the work can hardly be evaluated quantitatively. For the different methods of detecting planes from homographies a detailed evaluation is given in section 5.2.

5.1 Detection of Homographies

Detailed error analysis of the decomposition of image sequences into planes is difficult. First of all real video sequences do not provide a ground truth segmentation that could be used for numerical error analysis. But even more important such a decomposition into planar patches is not unique. Planar patches detected from sparse point correspondences are in fact typically smaller than the physical planes they represent, and finding the exact delineations of planar regions is a different issue not covered here.

We have performed experiments with different scenes and environments. In some rather artificial sequences, checkerboard calibration patterns were placed on a table and recorded with a handheld camera. The checkerboards provide high contrasts and sharp corners, that can be tracked well and provide good point correspondences over the image sequence. Another set of images was taken from publicly available sequences of architectural scenes showing model buildings. These kind of scenes are a typical application scenario for planar patch detection.

Example planes found with our algorithms are shown in Figure 2 for a calibration pattern scene and in Figure 3 for an architectural scene. Note the detected planes do represent planar image areas and correspond to physically present planes in the scene, no virtual planes are detected. As it was expected, the detected planes are typically smaller than the physical planes due to the sparseness of the point correspondences used to find them. Points assigned to a plane were not removed and therefore some planes are detected several times and do overlap. On the other hand this allows correct handling of points on the delineation of two planar patches. Note that point correspondences not lying in any of the planes are correctly identified, so if the observed objects are not planar, no false planes are detected.

In these example images, the planar patches are detected only and not kept from one frame to the next. Depending on the application, this temporary knowledge of coplanarity might be sufficient. Otherwise a homography tracking can be applied and simple methods to prevent overlapping planes from being detected over and over again could be thought of.

5.2 Detection of Cases Without Translation

In section 4 we have presented various ways of detecting camera motions without translational part. In these cases the homographies do not give us any information on coplanarity of points and hence no planes can be detected using the homographies.

To evaluate the performance of the individual methods, some video sequences with controlled camera motion were recorded. Mounted on a tripod, a camera captured a motion sequence with at least approximately a pure rotational motion. With a motorized zoom it was further possible to take influence on the intrinsic camera parameters without any other camera motion. So it was possible to acquire a ground truth classification of the camera motion and to compare the detected motion classes of “translation” and “no translation” with that ground truth.

In Figure 4 the different criteria from section 4 were compared for a sequence with pure camera rotation. The ground truth information is shown as a background coloring, where the white parts indicate no camera motion, yellow parts a camera rotation and green parts a sequence of images with non-zero camera translation. For each image frame the tests computed one value per detected homography, e.g. one ratio of eigenvalues. For the figure these values were averaged over several such tests (e.g. over the 5 planes detected in this frame). Note that due to constant and known intrinsic camera parameters, all criteria could be applied for the sequence with pure rotation. The short times with completely static camera were clearly identified by all criteria. The translational movement can also be clearly identified from the global homography criterion (line “global”). Also the singular value and eigenvalue criteria allow a classification of the camera movement, with some small false alarms around frame 45. The epipolar criterion seems to be severely affected by incorrect point matches however.

A similar comparison is shown in Figure 5 for variable intrinsic parameters, i.e. a zooming camera. Note we do not have accurate knowledge of the intrinsic parameters in this case and hence skip the singular value criterion. To allow comparison we did test the eigenvalue criterion however. It can be seen that the criterion incorrectly classifies the zooming camera, as expected. As described in section 4.2 the criterion needs constant intrinsic parameters to be valid. Both the epipolar and especially the global homography criteria allow a relatively good identification of the translational camera motion, however the results are far less clear compared to the sequence with a rotating camera.

Overall if the intrinsic calibration is known or constant, this knowledge should be used, as was seen in the test with pure camera rotation. In other cases the global homography criterion seems to perform sufficiently good as well. This was also confirmed in further qualitative tests with different sequences. The epipolar geometry most likely suffers from numerical instabilities and outliers of the point matching. Skipping the tests for a camera translation, one "plane" is detected covering all point correspondences in the image, unless a translational motion is present.

6 FURTHER WORK

The criterion derived from epipolar geometry currently does not provide a useful measure for the translational part, most likely due to the numerical instability of computing epipolar geometry. The normalized eight-point-algorithm used in this work already performs better than using unnormalized pixel coordinates, but still it is not robust against incorrect point matches. Using an improved algorithm could also render the epipolar geometry useful for homography estimation, as described in section 3.1.

Having found the coplanar point sets, the exact delineations of the planes are still unknown. A pixel-wise assignment of image points to physical planes is needed for various applications like exact scene representation or image based rendering. This can be solved with region growing algorithms, as was done e.g. in (Gorges et al., 2004) or with graph-cut related techniques. Both do need initial seed regions that can be generated robustly from the image data with our algorithms. And both have to be made aware of cases where it can not be inferred on coplanarity from homographies.

7 CONCLUSION

The aim of this work was to automatically detect planar features in image streams from handheld cameras. Various applications were mentioned in the introduction. In most of these a manual selection of planes is used. The few works dealing with the automatic detection of planes concentrated of finding image regions under homography. We have given a brief overview and presented a similar algorithm based on random sampling and iterative estimation of the dominant plane.

As we have shown, finding homographies between the frames of a sequence can not be enough for the detection of planes however. For camera movements without 3D translational part the common homography is not a sufficient criterion for the coplanarity of points. We have presented various methods to detect such cases and to prevent planes from being detected in case of no camera translation. These methods made use of known or constant intrinsic camera parameters or of the static-scene assumption, and hence can be applied to many different application scenarios.

In the experiments we have first shown that physically meaningful planes can be detected with the suggested approach. Also a comparison of the various methods for plane extraction from the homographies was given. Especially the cases of pure camera rotation and varying intrinsic parameters were investigated, exactly the cases where a homography does not contain information about the coplanarity of points. The sequences with a pure rotation could be identified clearly. It was more difficult to separate a change of intrinsic parameters from general camera motion. But using the appropriate methods it was possible as well.

REFERENCES

Baker, S., Szeliski, R. and Anandan, P., 1998. A layered approach to stereo reconstruction. In: Proc. Computer Vision and Pattern Recognition, Santa Barbara, CA, pp. 434–441.

Bartoli, A. and Sturm, P., 2003. Constrained structure and motion from multiple uncalibrated views of a piecewise planar scene. International Journal of Computer Vision 52(1), pp. 45–64.

Cobzas, D. and Sturm, P., 2005. 3d ssd tracking with estimated 3d planes. In: Proc. Second Canadian Conference on Computer and Robot Vision, Victoria, Canada, pp. 129–134.

Gorges, N., Hanheide, M., Christmas, W., Bauckhage, C., Sagerer, G. and Kittler, J., 2004. Mosaics from arbitrary stereo video sequences. In: Lecture Notes in Computer Science, Vol. 3175, 26th DAGM Symposium, Springer-Verlag, Heidelberg, Germany, pp. 342–349.

Hartley, R. and Zisserman, A., 2003. Multiple View Geometry in Computer Vision. 2nd edition edn, Camebridge University Press.

Odone, F., Fusiello, A. and Trucco, E., 2002. Layered representation of a video shot with mosaicing. Pattern Analysis and Applications 5(3), pp. 296–305.

Rother, C., 2003. Linear multi-view reconstruction of points, lines, planes and cameras using a reference plane. In: Proc. International Conference on Computer Vision 2003, Nice, France, pp. 1210–1217.

Schindler, K., 2003. Generalized use of homographies for piecewise planar reconstruction. In: Proc. 13th Scandinavian Conference on Image Analysis (SCIA2003), Springer-Verlag GmbH, Gotenborg, pp. 470–476.

Shi, J. and Tomasi, C., 1994. Good features to track. In: IEEE Conference on Computer Vision and Pattern Recognition CVPR, pp. 593–600.

Torr, P. H. S., Fitzgibbon, A. W. and Zisserman, A., 1999. The problem of degeneracy in structure and motion recovery from uncalibrated image sequences. International Journal of Computer Vision 32(1), pp. 27 – 44.

Triggs, B., 1998. Autocalibration from planar scenes. In: Proc. 5th European Conference on Computer Vision, Vol. 1, Springer-Verlag, pp. 89 – 105.

Zelnik-Manor, L. and Irani, M., 1999. Multi-view subspace constraints on homographies. In: Proc. 7th International Conference on Computer Vision, Vol. 2, IEEE Computer Society, Kerkyra, Corfu, Greece, pp. 710–715.