SENSOR POSE INFERENCE FROM AIRBORNE VIDEOS BY DECOMPOSING HOMOGRAPHY ESTIMATES

E. Michaelsen*, M. Kirchhof*, U. Stilla△

*FGAN-FOM Research Institute for Optronics and Pattern Recognition
Gutleuthausstrasse 1, 76275 Ettlingen, Germany
{michaelsen,kirchhof}@fom.fgan.de
△Photogrammetry and Remote Sensing, Technische Universitaet Muenchen
Arcisstrasse 21, 80333 Muenchen, Germany
Uwe.Stilla@bv.tum.de

Commission III, WG III/1

KEY WORDS: Geometry, Vision, Estimation, Navigation, Orientation, Infrared, Aerial, Video

ABSTRACT:

Airborne videos are gaining increasing importance. Video cameras are taking huge amounts of measurements for low costs. Their low weight and low requirement for energy makes them particularly attractive for small airborne carriers with low payload. Such carriers are discussed for military as well as for civil applications, e.g. traffic-surveillance. Often video cameras are used for documentation and reference in connection with other sensor systems. In addition to panchromatic or ordinary colour videos, nowadays also cameras operating in the thermal spectral domain gain attention. For the utilization of any stream of measurements taken from a moving platform the pose of the sensor in orientation and position has to be constantly determined. For airborne platforms often GPS and INS are used to acquire this information. However, the video stream itself provides also possibilities to estimate pose parameters. In this contribution we restrict our investigation to almost flat scenes but we allow oblique views both forward looking and side looking. The optical flow of the scene fixed structure on the world plane is estimated by a planar projective homography. This requires at least four point or line correspondences that can be traced over an appropriate number of frames. If the focal length is not changed and the camera has not been rotated, the proper transform will be restricted to a central collineation with five degrees of freedom. Two of these - giving the vertex or epipole - can be inferred directly from image correspondences. The remaining three are then estimated from the homography by solving a homogenous linear system. They give the axis or horizon, from which we obtain the rotational part of the pose, and a scale parameter for the speed to height ratio. Common level keeping flight manoeuvres where the epipole is close to the horizon lead to elations. Other manoeuvres - like e.g. landing - lead to homologies. The rotation-free calculations will also be appropriate if the camera rotation is known from another sensor. If the rotation between the frames is unknown the homography will be decomposed into a central collineation and an orthogonal rotation matrix. The five degrees of freedom of the collineation and the three degrees of freedom of the orthogonal rotation matrix sum up to eight, which is exactly the same number of degrees of freedom that a planar homography has. There is a set of analytic solutions to this equation system, of which the correct solution can be picked by heuristic considerations. We investigate the propagation of measurement errors through these calculations. Examples for such estimations are shown for thermal videos. Long focal lengths are unfavourable. The rotation-free decomposition gives more stability compared to the decomposition with rotation.

1. INTRODUCTION

1.1 Unmanned Aircraft and Airborne Video

New possibilities for a variety of tasks including traffic monitoring, disaster management, surveillance and military applications come with the increasing utilization of unmanned aircraft. These crafts can be built quite small and at low cost. The payload and power resources are limited, but almost always they will feature one or several digital video cameras. This contribution investigates a possible utilization of this sensor type for the pose estimation and thus navigation of the craft. Automatic control of unmanned aircraft by vision alone may be one goal while another one may be the combination of this information source with other sensors like inertial systems, laser range finders, altimeters, speed sensors or GPS. Here we only treat central perspective cameras that take the whole picture through one aperture at one time instance. Devices using analog video standard with two half-frames are included but need special care. Push-broom cameras and CCD-line scanners are excluded.

1.2 Properties of the Thermal Spectral Domain

For many tasks operability at any time of the day and also under bad weather conditions is desired. Electromagnetic waves in the thermal bands between 3µm and 5µm or between 8µm and 12µm give the opportunity to measure the black-body temperature radiation of the objects on the ground. The energy that is measured comes from emission rather than reflection. No external light source is needed. The transparency of the atmosphere in these two thermal bands is equal or better than in the visual band between 0.4µm and 0.8µm. For tasks like vehicle recognition or traffic surveillance thermal measurements give the unique opportunity to determine the operational status of objects. Running engines emit thermal radiation. Today, the radiometric resolution and dynamic range

is usually quite good, contemporary cameras often give 16 bit data with twelve bit information.
There are also disadvantages for such cameras: Usually they are much more expensive than standard digital CCD video cameras. If we take diffraction at the aperture as limit for the angular resolution a lens for a thermal camera may have to be ten times bigger than the equivalent lens for the visual camera. Also often the detector has to be cooled down to very low temperatures. Therefore thermal cameras are usually bigger and need more energy than visual cameras. They also do not give any spectral measurements like a colour CCD camera does. Some modern thermal cameras have a focal plane array sensor but some systems still have only a small number of sensors. These cameras compose the image using moving mirror systems, which gives special distortions in the image geometry.
Because of the lack of colour and because of frequent appearance of non-structured homogenous regions with no temperature differences thermal videos pose a more difficult challenge to geometric estimation procedures. Therefore all our examples are picked from this domain. The algorithms also work for aerial videos of the visual spectral domain with colour being an important feature for correspondence assessment.

2. ESTIMATING POSE FROM HOMOGRAPHIES

2.1 Interest Point Locations

It is not possible to localize correspondence between different frames if the object is homogenous in that location. If an edge or line structure is present at a location in the 2-d image array there may still be an aperture problem. Secure point correspondence can only be obtained at locations where a corner, crossing or spot is present. It is proposed to use the averaged tensor product of the gradient of the grey-values (Förstner, 1994). Interest locations are given where both eigen-values of this matrix are non-zero.

2.2 Assessment of Correspondence

Correspondence between locations in different frames of a video can be assessed using grey value correlation. Still there may be problems with repetitive structures. However, there will usually be a prior estimate for a location correspondence. Then this might be used to assign regions of interest in one image to interest locations in the other image or to form the overall assessment as product of correlation and prior probability. Regions of interest or the variance of priors will be quite narrow for immediately successive frames of the video.

2.3 Planar Homographies

Given a perspective projection and a rigid planar scene the movement of locations in the image is determined by a planar homography $x' \approx Hx$, where the image location correspondence (x,x') is written in homogenous coordinates, H is a 3x3-matrix and $\approx$ means equality up to an unknown scale factor. Given a set of at least four correspondences H can be estimated using direct linear transformation (DLT) (Hartley; Zisserman, 2000).
We assume the inner camera parameters to be known and the 2-d coordinates to be normalized such that the focal length equals one and the origin is at the principal point. Shifting the principal point has no major impact on the precision of the estimations. But the influence of the focal length is considerable: Either we may do the estimation first and transform to normalized coordinates afterwards using

$$\begin{pmatrix} h_{11} & h_{12} & h_{13} \\ h_{21} & h_{22} & h_{32} \\ h_{31} & h_{32} & h_{33} \end{pmatrix} \xrightarrow{f} \begin{pmatrix} h_{11} & h_{12} & 1/f\,h_{13} \\ h_{21} & h_{22} & 1/f\,h_{32} \\ f\,h_{31} & f\,h_{32} & h_{33} \end{pmatrix} \quad (1)$$

Then we will enlarge errors on the projective elements h_{31} and h_{32} respectively by factor f and diminish errors on the translation elements h_{13} and h_{23} with the same factor.
Or we may do the transform on the image coordinates x and x' dividing the first two components of them through f and go into the DLT system with these smaller entries. This has a similar effect: The equation system will not be balanced. Entrances responsible for unknown variables h_{11} h_{12} h_{21} h_{22} in the affine section will be smaller than the entries for the unknown translation elements h_{12} and h_{13} with approximately the same factor f and for the unknown projective elements h_{31} and h_{32} respectively there will be very small entrances (factor f^2).

2.4 Decomposition of Homographies

Given an estimate for the normalized planar homography H we can reconstruct the pose parameters using the decomposition

$$H = R - tn^T, \quad (2)$$

where R is an orthogonal rotation matrix, t is the translation of the camera and n is the surface normal (Faugeras, 1995). This representation sets the origin of the 3-d system into the centre of the second camera. R contains three degrees of freedom that may be extracted as successive rotation angles or as normalized axis in 3-d and turning angle around it. The vectors n and t together contain five degrees of freedom because n will be normalized setting the distance of the second camera to the plane to one, while t is a 3-d translation.
The absolute scale cannot be determined from the image sequence alone. This requires additional information e.g. from an altimeter or from a speed sensor.
In rural areas the plane will be a good approximation for the ground plane. In urban areas most visible structure will result from the roofs. So the plane will be at average roof height over ground. The vector n will still be a good approximation to zenith direction. We will not get information on the north-direction from the images unless we rely on shadow and daytime analysis. There will be no geo-reference from the images as long as we have not recognized or matched objects from the images to map objects.
We assume sufficient movement of the air-craft. This is important, because the decomposition of homographies needs to distinguish the translation-free case from mappings with translated cameras.

The rotation free case: Often the camera will be mounted on a stabilized platform or the camera rotation will be measured by an inertial device giving much more precision than the estimation from the camera may yield. This known rotation may be applied as homography to the coordinates of the first image and then we may assume R to be the identity. Then the homography is restricted to be a central collineation with real eigenvalues, which is either a planar homology or elation (Beutelsbacher; Rosenbaum, 1998). Considering the homology case first we may scale H such that the double eigenvalue equals one. The corresponding 2-d eigenspace is the horizon line. This is a straight line of fixed-points (the image of the intersection of the plane n with the plane at infinity) and n also gives its Hessian normal form. The other eigenspace is 1-d and gives the epipole and translation t. The eigenvalue

corresponding to this eigenvector is $1\text{-}t^T n$. And since n is normalized we get the proper length of t from this equation. This solution is unique up to change of sign of n and t.
The eigenvalue calculation of a homography estimated from correspondences may also result in a pair of conjugated complex eigenvalues and a single real eigenvalue. In the rotation free case such result cannot be used for pose estimation. A homology has to be searched for that is closest to the estimated homography.

The elation case: The rotation free homography becomes an elation if the epipole lies on the horizon line i.e. $t^T n=0$. This is not an exception but common for many standard flight manoeuvres (keeping level). Such mappings have a triple eigenvalue with a corresponding eigen-space of rank two (the horizon line). The epipole cannot be stably estimated from the eigen-spaces of the homography. Instead it can be estimated from pairs of correspondences directly by intersecting the correspondence straight lines. We used the correspondence-pairs that are part of the best solution for homography estimation (see Sect. 3) and iterative re-weighting to minimize the influence of outliers in this estimation. Given an epipole estimation t and setting the Rotation R to identity I equation (2) becomes linear in the plane parameters n. To cope for the unknown scaling of t we set a forth scalar parameter μ and get

$$H = I - \mu t n^T . \tag{3}$$

Dividing this equation by μ we get a set of nine homogenous equations in the four unknowns n and $1/\mu$. It is solved by singular value decomposition. In fact this method is applicable for all central colineations, i.e. not only for elations but also for homologies. It may replace the eigen-space construction as well.

With rotation: Taking non-trivial rotations R into account the homography is transformed into diagonal form using singular value decomposition $H=UH'V$ with orthogonal rotations U and V. Equation (2) is transformed to $H'=R'-t'n'^T$ where $R=UR'V^T$, $t=Ut'$ and $n=Vn'$. This new equation can be solved analytically. Assuming the singular values h_1, h_2 and h_3 in the diagonal matrix H' to be sorted and h_2 scaled to one the rotation may be restricted to the Y-axis and the Y-components of t and n set to zero:

$$\begin{pmatrix} h_1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & h_3 \end{pmatrix} = \begin{pmatrix} \cos\beta & 0 & -\sin\beta \\ 0 & 1 & 0 \\ \sin\beta & 0 & \cos\beta \end{pmatrix} - \begin{pmatrix} t'_X n'_X & 0 & t'_X n'_Z \\ 0 & 0 & 0 \\ t'_Z n'_X & 0 & t'_Z n'_Z \end{pmatrix} \tag{4}$$

This leads to

$$n' = \begin{pmatrix} s_1\sqrt{\dfrac{h_1^2-1}{h_1^2-h_3^2}} \\ 0 \\ s_2\sqrt{\dfrac{1-h_3^2}{h_1^2-h_3^2}} \end{pmatrix}, \; t' = (h_1-h_3)\begin{pmatrix} n'_X \\ 0 \\ -n'_Z \end{pmatrix}, \; \sin\beta = (h_1-h_3)n'_X n'_Z \tag{5}$$

where all four combinations of signs $s_1=\pm 1$ and $s_2=\pm 1$ are permitted. Transforming these solutions back to equation (2) we obtain four solution sets for R, t and n.
A critical situation occurs where the solutions branch, i.e. where the value in one of the roots becomes small or two of the singular values are nearly equal. This is the case if t and n are parallel i.e. the craft is directly going nadir – an unusual flight-manoeuvre. The method will completely break down for three equal singular values i.e. $t=0$. We exclude this case because it is physically impossible for aircrafts.

Special forms: The different applications lead to different structures of the homography matrices. Equation (6) lists some important cases. A camera looking directly nadir down to the surface and the craft moving in X-direction will produce something close to the form H_n. Side-looking obliquely mounted cameras will give almost H_s. A forward-looking geometry will result in something close to H_f.

$$H_n = \begin{pmatrix} 1 & 0 & -u \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}, \; H_s = \begin{pmatrix} 1 & -uv & -u \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}, \; H_f = \begin{pmatrix} \dfrac{1}{1-u} & 0 & 0 \\ 0 & \dfrac{1}{1-u} & 0 \\ \dfrac{-uv}{1-u} & 0 & 1 \end{pmatrix} \tag{6}$$

u is a velocity parameter and v is a parameter for the tilt of the plane. Note that H_s is an example for the elation case discussed above.

Error Propagation: These frequent special forms are used to propagate small displacement errors in the correspondences into the estimated pose parameters. For this purpose we use the following set of four correspondences

$$\{P_1,P_2,P_3,P_4\} = \left\{\begin{pmatrix} 1 \\ -1 \\ 1 \end{pmatrix}, \begin{pmatrix} -1 \\ 1 \\ 1 \end{pmatrix}, \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}, \begin{pmatrix} -1 \\ -1 \\ 1 \end{pmatrix}\right\} \tag{8}$$

We set motion to $u=0.01$ for the matrix H_s and H_n (H_s for a angle of 45°) and computed corresponding points from this. Then we put an error ε to the last point computed the disordered homography and decomposed it again. Displacement results for the vector t are listed in Table 1.

	ε1=0.0005	ε2=0.001	ε3=0.0025
Rotation-included			
f=10	11%	23%	61%
f=100	166%	281%	665%
Rotation-free			
f=10	1.7%	3.2%	7.4%
f=100	1.3%	2.5%	12%

Table 1. Sensitivity of translation t to errors at different focal lengths. Deviations are given in ratio to the length of t.

These are fairly small errors (ε3 being some half pixel). They can only be reached by using more than four correspondences and a robust estimation method. Also a ratio of 10 or 100 of the focal length to half of the image size is common for IR-cameras.

2.5 Non-projective and projective Distortions

Particularly thermal IR cameras of older construction type often show strong non-projective distortions. They only have a small number of sensors that are used to scan the image successively. The rotating mirrors that are used to map the image to the sensors cause a non projective mapping. Examples for such data are Videos I and II from Sect. 4. If the construction details of camera are not known the non-projective part of the distortion

may be estimated by a non-linear function $x_e=q(x_n)$ where x_n is the distorted location in the image and x_e the location used for estimation of the homography H_e. We used a cubic 2D-polynome for q and estimated the parameters by forcing locations that are known to be collinear in the scene to be also collinear in the images. The effect of this function q is visualized in Figure 1.

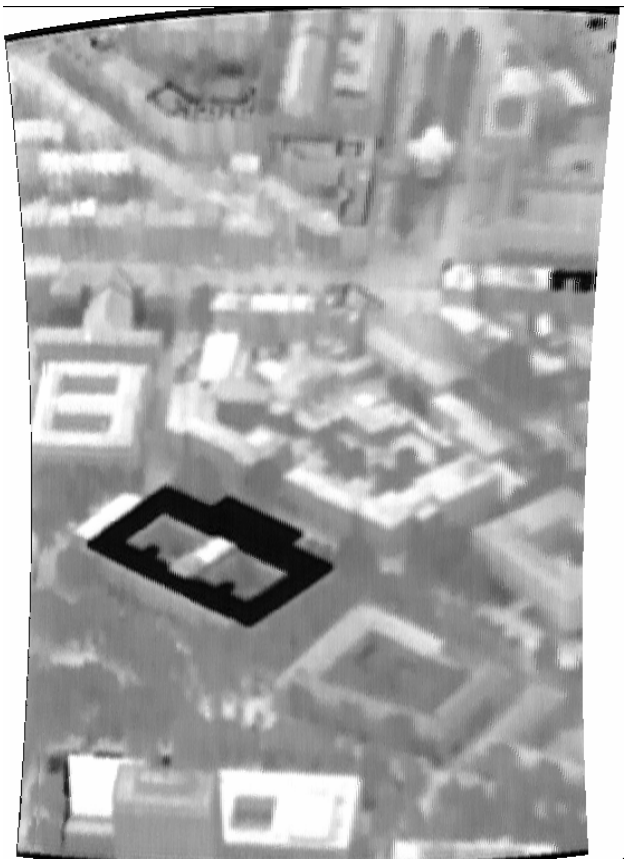


Figure 1. Example for the cubic distortion correction applied to video VideoII; these images are usually not calculated, only the positions of the interest points are corrected.

Calculating pose from such homographies H_e estimated from correspondences (x_e, x'_e) directly may lead to systematic errors because the camera distortion (and its non-linear correction) may also contain an unknown projective part H_d. For such distortions collinearity is an invariant. H_d applies to both positions of each correspondence: $H_d\, x'_e = H\, H_d\, x_e$. If there is a ground-truth homography G for some images of the scene the equation $H_e=H^{-1}_d\, G\, H_d$ may be transformed to $H_d\, H_e\, -G\, H_d = 0$, and used to estimate H_d. But care has to be taken that H_d is chosen with a sufficient determinant. It should be chosen from a particular family of mappings like shearing-mapping, rotation or projective distortions. Systematic errors for a particular set-up can be measured if ground truth is given with a calibration run.

Many contemporary thermal cameras feature focal plane array sensors and can thus be handled like ordinary CCD TV-cameras: For normal or long focal lengths pin-hole camera models will usually be precise enough and the distortion of wide angel lens may sufficiently well be treated by a single quadratic term. An example for data from such a modern device is Video III in Sect. 4.

3. SEARCHING FOR PROPER CORRESPONDENCE SETS WITH PRODUCTION SYSTEMS

A planar homography $h(p)=Hp$ can be linearly calculated from four correspondences c=(p′,p) by direct linear transformation (DLT) (Hartley; Zisserman, 2000), but they must not be in special configurations (like if three of the four points are collinear). If there are more than four correspondences available a squared residual error sum R is minimized.

$$R = \sum \varepsilon(p', Hp)^2$$

There is again a DLT solution to this, where ε is an algebraic error that approximates the inhomogeneous 2-d error provided the point coordinates are given in proper normalization (Hartley; Zisserman, 2000). However, this process being a least squares method is very sensitive to the inclusion of outliers in the calculation. Therefore a robust estimation method is required that can detect and eliminate the false correspondences.

3.1 Robust Estimation

Several proposals have been made to minimize the influence of such gross errors. One example is the iterative re-weighting approach (Holland; Welsch, 1977). This method avoids hard decisions on the set of measurements. However, the most popular approach today is the well known random sample consensus (RANSAC) approach (Fischler; Bolles, 1981).

The Random Sample Consensus Method: Let $C=\{c_1,...,c_n\}$ be the set of correspondence cues obtained from the images. It is expected that C is the disjoint union of a set of correct correspondences C_c and outliers C_o. The goal is to identify these sets by automatic means and to minimize the goal function

$$H_{opt} = \arg\min_{H}(R(H, C_c))$$

This minimization varying the homogenous matrix H while keeping C_c fixed is a straight forward linear computation using DLT. However, searching C for the proper subset C_c poses an exponential challenge (in n). The RANSAC-proposal recommends probing the power-set of C by drawing random minimal subsets. Here these are quadruples $s=\{i_1,...,i_4\}$ from $\{1,...,n\}$. Each of these samples leads to a hypothesis for H_s (at least if the calculation succeeds). And for every such hypothesis the residual error for all elements in C is determined.

$$r_{s,i} = \varepsilon(p'_i, H_s p_i)^2 .$$

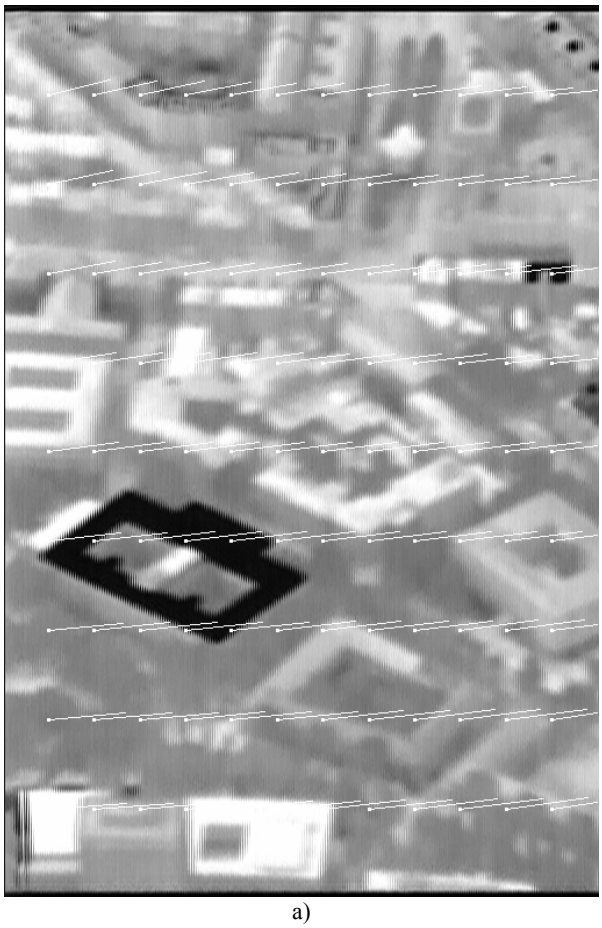
Using a global threshold δ the consensus set of the sample is defined as $\{c_i \text{ in } C: r_{s,i} < \delta \}$. The largest consensus set is supposed to be a good approximation for C_c. Usually it is too time consuming to check all $\binom{n}{4}$ samples. There are decision theoretic considerations that give hints on how many samples should be drawn given an expected outlier-rate, a variance for the positioning of correct correspondences and a significance level [Hartley]. It is also possible to continue probing until a predefined minimal consensus is reached, or – in an any-time manner – until a solution is demanded by exterior time constraints.

If the set C is not equally distributed in the image the method will adapt the transform with more weight on densely populated regions. An isolated important correct correspondence in an otherwise homogenous image region may either end up as "outlier" or it will have equal weight like any other single correspondence in the calculation.

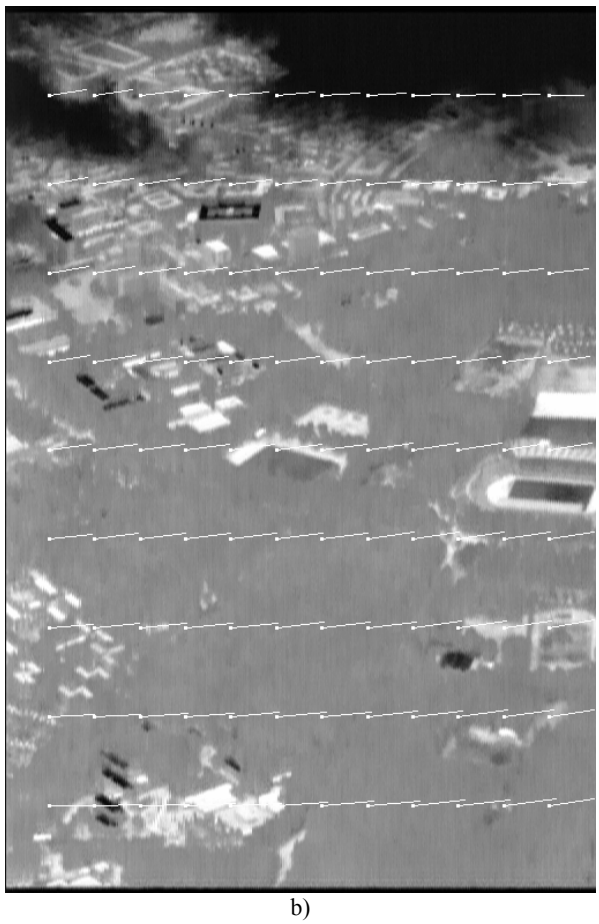
Good Sample Consensus: Already in (Fischler; Bolles, 1972) an improvement of the RANSAC paradigm by replacing the random samples by samples that are drawn according to an assessment criterion is sketched. Following this idea we implemented the following approach:

1. Locations are picked from each image which contain enough structure to allow a correspondence test with high significance (Foerstner, 1994). More significant locations gain higher priority.
2. Each sample c_i is evaluated according to its correlation. Samples with high evaluation gain high priority.
3. Pairs of correspondences (c_1,c_2) are formed and assessed according to their Euclidian distance. Correspondences that are far apart gain high priority.
4. Two pairs form a quadruple $(s_1,\ldots,s_4)$ of correspondences. It is assessed according to the area covered by the smallest of the four triangles formed by the points in one of the images. This property will be zero if three of the points are collinear. A quad with large minimal area gets high priority.
5. Each quadruple defines a homography using DLT. In the space of homographies a metric is defined and quadruples that vote for close transforms are merged. The parameters of such a cluster of homographies are recalculated using the version of DLT that minimizes the residual error R for all correspondences preceding it. We call this correspondence set the consensus of the cluster. It is assessed according to the size of the consensus and also to the assessments of its members and according to geometric properties like the size of its convex hull.

Good Sample Consensus method has been motivated and discussed in (Michaelsen; Stilla, 2003). The constructions and assessments are coded as productions and entered into a production system. The production system is run on the data using a data-driven bottom-up control that has any-time capabilities (Stilla, 95).



a)



b)

4. EXPERIMENTS AND CONCLUSION

4.1 Experiments with Aerial Thermal Videos

Three video sequences taken from helicopters or aircrafts have been used to verify the error behaviour of homography estimation and pose estimation using decompositions of such homographies. All are taken in the thermal spectral domain. Fig. 2 shows example frames for each video.

Video I: Oblique side-looking sequence on urban region in the city of Karlsruhe (buildings on flat terrain) taken with a TICAM camera from an airplane flying at approximately 3000m height. Such cameras give strong non-projective distortions. The camera was zoomed to 540mm focal length. Detector spacing in x- and y-direction is 50µm. So this is an extreme tele-lens perspective.

Video II: Oblique side-looking sequence on the same urban region (including a lot of homogenous park area) taken with the same TICAM camera from an airplane flying at approximately 3000m height. The camera was zoomed to 212mm focal length still giving a small field of view.

Video III: Forward-looking sequence from a rural region with a little creek (trees, bushes, etc.) taken with a focal plane array camera from a helicopter flying at very low altitude. Such cameras give almost no non-projective distortions. The camera has fixed standard field of view. The focal length to detector spacing ratio is approximately the same as for Video II.

c)

Figure 2. Example frames from thermal image videos with homography estimation overlaid: a) Video I, oblique with very small field of view; b) Video II, oblique with still small field of view; c) Video III, forward looking, low flying and strong rotations.

For Video I there was a ground-truth file containing pose estimates obtained by a priori GIS/INS recordings and posterior back-section with geo-referenced building models. This enabled the estimation of the projective distortion matrix and systematic offset like outlined in Sect. 2.5. The same distortion parameters were used for Video II. Video III was regarded as distortion-free (but with camera rotations). The outlier-rate and the standard deviations for inliers where estimated. Deviations are given in relation to the length of the measured vectors. Outliers are usually defined as having more than 100% deviation (except Video II with rotation, were 500% were used for *t*).

	Video I	Video II	Video III
Outlier-rate Rotation-included	100%	71%	32%
Deviation t	-------	385%	43%
Deviation n	-------	64%	41%
Deviation R	-------	5%	10%
Outlier-rate Rotation-free	58%	16%	31%
Deviation t	78%	38%	56%
Deviation n	69%	39%	48%

Table 2. Deviations and outlier-rates

4.2 Conclusion

Particularly, if long focal lengths are used the estimation accuracy for the camera pose that can be achieved through full homography decomposition may be rather poor. With Video I it turned out a complete random number generator. Often INS-sensors are mounted on the same platform anyway that will give high precision. In these cases a rotation-free decomposition is favoured with the homography being a homology or an elation. The elation case is not exception because aircrafts are often operated at level that means with the epipole on the horizon. Therefore the eigen-space decomposition is not recommended. Instead the rotation free case allows estimating the epipole directly from the best sub-set of correspondences. Subsequently the plane parameters can be inferred linearly and unambiguously from the decomposition. Pose estimation by decomposing the homography of the image flow measures velocities over ground in relation to flight altitude and the absolute nadir direction (at least if the scene plane is levelled). It may complete other sensors like INS giving accelerations and rotations, GPS giving absolute locations, speed sensors giving speed in air, altimeters giving absolute height and magnetic compasses giving geographic direction. Special care has to be taken for non-projective distortions of the camera and lens, particularly with thermal cameras of the non-focal-plane-array type. Such distortions may well be misunderstood by the decomposition as rotations. It is desirable to calibrate a systematic offset from comparing the system to ground-truth.

References

Ben-Ezra, M., Peleg, S., Werman, M., 1999. *Real Time Motion Analysis with Linear Programming.* CVIU, Vol.78, pp. 32-52.

Beutelsbacher, A., Rosenbaum, U., 1998. *Projective Geometry*. Cambridge Univ. Press, Cambridge.

Faugeras, O., 1993. *Three-Dimensional Computer Vision*. MIT Press, Cambridge, Mass,.

Fischler, M. A., Bolles, R. C., 1981. Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography. Comm. Assoc. Comp. Mach., Vol. 24, pp. 381-395.

Foerstner, W., 1994. *A Framework for Low Level Feature Extraction.* In: Eklundh J.-O. (ed). *Computer Vision – ECCV 94*. Vol. II, B1, pp. 383-394.

Hartley, R., Zisserman A., 2000. *Multiple View Geometry in Computer Vision*. Proc. Cambridge University Press, Cambridge.

Holland, P. W., Welsch, R. E., 1977. *Robust regression using iteratively reweighted least-squares.* Comm. Statist. Theor. Meth., Vol. 6, pp. 813-827.

Jurie, F., Dhome, M., 2002. *Real Time Robust Template Matching*. BMVC-2002, pp. 123-132.

Michaelsen E., Stilla U., 2003. *Estimation of vehicle movement in urban areas from thermal video sequences.* 2nd GRSS/ISPRS Joint Workshop on Remote Sensing and data fusion on urban areas, URBAN 2003, IEEE, pp 105-109.

Michaelsen E., Stilla U., 2003. Good sample consensus estimation of 2d-homographies for vehicle movment detection from thermal videos. In: Ebner H, Heipke C, Mayer H, Pakzad K (eds) Photogrammetric Image Analysis PIA'03. International Archives of Photogrammetry and Remote Sensing. Vol. 34, Part 3/W8, pp 125-130.

Torr P. H. S., Davidson C., 2003. *IIMSAC: Synthesis of importance sampling and random sample consensus.* IEEE – PAMI, Vol. 25(3), pp. 354-364.

Sawhney, H. S., 1994. *Simplifying motion and structure analysis using planar parallax and image warping.* ICPR 94, IEEE-Press, Los Alamitos, Ca., Vol. I, pp. 403-407.

Stilla, U., 1995. *Map-aided structural analysis of aerial images.* ISPRS Journal of photogrammetry and remote sensing, Vol. 50(4), pp. 3-10