GOOD SAMPLE CONSENSUS ESTIMATION OF 2D-HOMOGRAPHIES FOR VEHICLE MOVEMENT DETECTION FROM THERMAL VIDEOS

Eckart Michaelsen, Uwe Stilla

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

Commission III

KEY WORDS: Infrared surveillance, Machine vision, Object recognition, Urban areas, Vehicles, Robust estimation

ABSTRACT:

In this contribution we describe a method to assess the activity of vehicles based on airborne image sequences taken by an infrared camera. Active vehicles often appear as a configuration of a dark and a bright spot close to each other. The sensor movement is inferred from image sequences. Due to the fast velocity of the platform estimations of vehicle movements require a precise measurement of the sensor movement. The camera may be tilted with the aircraft giving arbitrarily oblique views. Camera movements are treated as projective 2D-homographies. For the search of a subset of image point correspondences that is free of outliers and gives a precise estimate of the movement we use a production system implementing good sample consensus (GSAC). This new method is derived from the well known RANSAC-decisions and improves them by preferring good samples to random samples. As assessment criterion for minimal samples the area of the smallest triangle in the sample is used. We motivate the criterion for the quality of samples by error propagation through the estimated homography. A comparison is made with other robust estimation techniques namly RANSAC and iterative re-weighted least squares.

1. INTRODUCTION

1.1 Vehicle detection

Vehicle detection has been an important topic in computer vision for a long time (Dreschler & Nagel, 1982). Optical flow helps a lot in segmenting the moving vehicle from the stationary background. Such successful approach is still being pursued, even if a geometric model of the vehicle is utilized (Haag & Nagel, 1999). If not only the vehicle is moving, but also the camera, the advantage of using videos and a simple threshold on the optical flow will disappear. But a scene fixed camera will only capture the activity in a certain very limited area. An airborne camera is much more flexible and can cover large areas. Vehicle detection from airborne videos has also been addressed by Partsinevelos et al. (2000).

Apart from movement temperature is another important cue to active vehicles. Furthermore, thermal images provide the opportunity to reveal the activity in an urban area by day and night. We propose to use an aircraft with a thermal camera for estimating the vehicle activity in urban terrain. The appearance of vehicles with this sensor depends on many factors, e.g. the daytime and the engine temperature. Passive vehicles often appear as single spots darker than the surroundings. They appear grouped into rows along the margins of roads or in parking lots (Michaelsen & Stilla, 2001). Active vehicles often appear as a pair of spots. This configuration consists of a bright spot resulting from the engine and a darker spot close to it resulting from the rest of the vehicle. In urban areas other objects may have the same property. But, the evidence for a car will be high, if such a pair of spots is moving along in the scene in the direction given by the spot pair.

1.2 Automatic Estimation

It is desirable to do the extraction, recognition and estimation without human interaction. This opens the way to do a large portion of the work on the fly in the aircraft. If we do not transmit all the images but only the estimations of what we want to investigate, we spare a lot of transmitting channel capacity. On the other hand this approach needs prior attention to robustness. Users will only trust in the results of automatic procedures if they are convinced of the robustness and precision of the outcome. Major sources of breakdown are to our experience the presence of un-modelled objects and clutter in the scene and erroneous correspondences between images due to partial occlusion. However, changes of lighting which are a problem in the visual spectral domain are not so serious in the thermal domain. Estimating 2D-projective homographies from grey-value templates or other features is rather instable. Lately, however, there have been proposed new estimation methods that yield a remarkable robustness (Jurie & Dhome, 2002). This paper makes a new proposal in this direction.

2. COMPARING EXISTING ROBUST ESTIMATION METHODS

The choice of an estimation technique preliminarily has to decide which kind of error function is to be minimized. Then the algorithmic approach for the given minimization problem is chosen.

2.1 Error Functions for Homography Estimation

Although we treat only 2D-projective homographies here, the assertions on the error function to be minimized easily generalize to problems like the estimation of similarities, fundamental matrices and trifocal tensors.

One straight forward possibility for an error function is the sum of absolute errors. Unfortunately, Euclidian point to point distance is not linear and not even differentiable. But there is, e.g. the point to line distance. Ben-Ezra et al. (1999) see an advantage in using such error function. Their approach leads to systems of linear in-equations that can be handled by simplex algorithms. Compared to quadratic errors, linear errors are less sensitive to outliers.

Another possibility for the choice of the goal function is the sum of the squared errors. Since the days of Gauss this has become the scientific standard approach known as LMSE (least mean square error). It has proven very useful in many domains. Moreover, LMSE will be the only correct optimal choice if the distribution of the errors is assumed to be normal. The sum of squares of the errors is a continuous and differentiable entity. Its analytic handling usually leads to linear equation systems that may be solved by standard techniques. For the remainder of this contribution we will therefore follow this line.

2.2 Robustness in the Presence of Outliers

Robust estimation of entities like homographies, fundamental matrices and trifocal tensors from sample correspondences has gained considerable attention. If outliers (erroneous correspondences) can not be avoided, a simple least square error approach will suffer severely even from a small percentage of outliers. Due to the square in the error, an extreme outlier will have an enormous impact on the calculations.

A straight forward approach to avoiding these difficulties is iterative re-weighting least squares (IRLS - Holland & Welsch, 1977). To this end the inverse of the residual of the least squares solution of each correspondence of the complete sample is used to re-weight the influence of it. Correspondences yielding a large residual error will be punished and correspondences yielding a small error will gain more influence. If a large portion of the correspondences is expected to be wrong, this may lead to local minima. The convergence of IRLS to the desired minimum is theoretically not guaranteed. It may end up with zero-error and thus infinite weight on an arbitrary minimal sample (quadruple with homographies) and random small weights on all other members. However, in practice we found that it does converge slowly but robustly to good solutions. IRLS-estimation of 2D-homographies is available in publicly code libraries (ISBE, 2003). Proposals are made how to handle occlusion outliers and lighting changes within the IRLS-method (Jurie & Dhome 2002). The main advantage of IRLS is the avoidance of decisions.

Another approach frequently found is the random sample consensus method (RANSAC) (Fischler & Bolles, 1981). To this end the calculation is performed on small – often on minimal – samples. These are picked at random from the complete sample of correspondences. The result of the calculation is tested on all the other correspondences giving a residual error. If this error is sufficiently small, the correspondence will be termed to be in consensus with the actual sample. After repeating this procedure for a predetermined number of such samples the search is terminated. The sample with the highest consensus is chosen and the corresponding consensus set is used to determine the estimation by mean squared error minimization. The disadvantage of performing early decisions is diminished by the use of decision theory for determining the threshold from statistics of relevant data and by using multiple decisions and the consensus principle. There is an elaborated theory for the choice of the two parameters (number of samples and threshold) from the usual portion of outliers, a standard deviation of the error of the position of inliers and a significance level (Hartley & Zisserman, 2000).

3. GSAC: A NEW STRATEGY FOR ROBUST ESTIMATION OF HOMOGRAPHIES

The quality of the estimation of homographies depends on the mutual positioning of the corresponding features. For improving the quality of the estimate we suggest to use good samples of corresponding features instead of random samples. We call this strategy "Good SAmple Consensus (GSAC)". To this end an assessment for samples has to be defined.

3.1 Motivation of GSAC by Considering Error Propagation

For airborne detection of vehicle movement in urban areas camera orientations close to nadir are used. For this situation we will not loose generality if we assume the true 2d homography H to be the identity. Let us assume a configuration of five points that are given as

$$P_0 = \begin{pmatrix} 0 \\ -a \end{pmatrix}, P_1 = \begin{pmatrix} 1 \\ 0 \end{pmatrix}, P_2 = \begin{pmatrix} 0 \\ 1 \end{pmatrix}, P_3 = \begin{pmatrix} -1 \\ 0 \end{pmatrix}, P_4 = \begin{pmatrix} 0 \\ -1 \end{pmatrix}$$

where $0 \leq a < 1$ is a parameter. The minimal sample size of four point correspondences is used. There are five possibilities to draw such a sample from the configuration characterized by the element that is not contained in the sample. One of the four sample points is assumed to be disturbed by a small error ε (see Fig. 1).

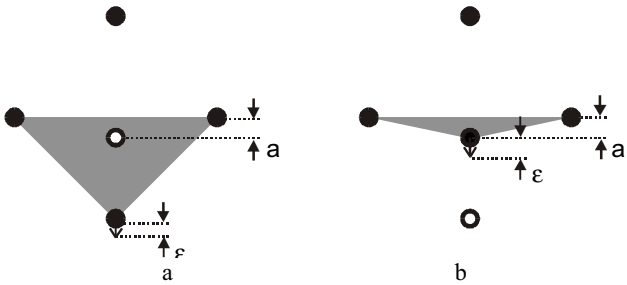

Figure 1. Quality of samples. a) benign configuration, b) unstable configuration. The sample is drawn as solid dots, while the test point is indicated as empty dot.

The influence of this error on the estimation of the homography H_ε and on the displacement it causes for the test point is calculated. As test point we use the point that is not contained

in the sample. The first sample (Fig 1a) consists of the points $P_1,\ldots,P_4$ (leaving out P_0). We disturb the corresponding point for P_4 in its second component by subtracting ε and get the estimation

$$H_e = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1+\delta & 0 \\ 0 & \delta & 1 \end{pmatrix}, \quad \text{where} \quad \delta = \frac{\varepsilon}{2+\varepsilon}. \tag{1}$$

This estimate is tested with the identity on P_0 setting $a_0=0$, $a_1=0.1$ and $a_2=0.5$. Assuming ε=0.001 we get for a_0 a displacement error of zero (the x-axis is a fix point straight of H_e). The displacement error for a_1 is approx. 0.000054 and for a_2 approx. 0.00037. We may average the squared displacement σ within the margins of our picture (e.g. x,y ∈ [-1,1]) using Eq. 2 and obtain a very small value σ≈0.00000016. We ascertain that such sample configuration is benign.

$$\sigma = \frac{1}{4} \iint_{x,y\in[-1,1]} \left[\left(\frac{x}{\delta y+1} - x \right)^2 + \left(\frac{(1+\delta)y}{\delta y+1} - y \right)^2 \right] dx\, dy \tag{2}$$

The second sample (Fig. 1b) consists of the points $P_0,\ldots,P_3$. (leaving out P_4). Now we disturb the corresponding point for P_0 in its second component by subtracting ε. This gives a homography estimation of the same matrix form

$$H_e = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1+\delta & 0 \\ 0 & \delta & 1 \end{pmatrix}, \quad \text{but} \quad \delta = \frac{\varepsilon}{a + a^2 + a\varepsilon}. \tag{3}$$

For $a_0=0$ this will obviously not work. In fact the equation system will have a defect if three points of the sample are collinear. In this case the solution will not be unique. We test displacements with the identity on P_4 for the other settings of parameter a. Setting $a_1=0.1$ gives a displacement error of approx. 0.018 and for $a_2=0.5$ approx. 0.0027. That means, configurations with one point close to a straight through two other points should be avoided.

Also for this σ we may evaluate the integration within the image using Eq. 2. We obtain much larger mean squared errors than for the benign setting, namely σ≈0.00000115 for $a_2=0.5$ and σ≈0.00005376 for $a_1=0.1$. The latter is more than 300 times bigger than with the benign setting. Even if we consider that this is a squared entity and take the root, there will still remain a factor of approx 18 for the standard deviation.

Pure RANSAC will treat all samples equal. Two samples (leaving out P_1 or P_3) contain collinear triples and thus lead to a defect in the equation system. Two samples (leaving out P_2 or P_4) are instable. We did not treat the sample leaving out P_2 but it is similar to the one leaving out P_4. At least for small settings of the configuration parameter (e.g. $a_1=0.1$) the test point will be falsely rejected from the consensus set, because the residuum is 18 times larger then σ. Only the fifth possibility (leaving out P_0) leads to the correct consensus set of all five points. If the random choice of a RANSAC run happens to contain this sample, it will succeed. Otherwise it will fail.

In the GSAC strategy the samples are assessed according to the value q which is proportional to the area covert by the smallest of the four triangles contained. It can be calculated from the difference vectors $(d_{i,x}, d_{i,y})^T = P_i - P_{(i-1) \bmod 4}$ using

$$q = \min_{i=0,..,3} \left| d_{i,x} d_{(i+1) \bmod 4,y} - d_{(i+1) \bmod 4,x} d_{i,y} \right| \tag{4}$$

These triangles are shaded grey in Fig. 1. For the first sample leaving out P_0 all four triangles have equal content one. The instable samples leaving out P_4 or P_2 contain the triangle $P_0P_1P_3$ which has area a. For a small value of parameter a this will give a bad assessment. This assessment criterion also prefers samples that cover large areas of the image. The samples leaving out P_3 or P_1 contain collinear triples. They will gain assessment q=0. Finding the correct consensus is guaranteed.

3.2 Implementing GSAC in a Production System

The GSAC strategy is implemented in a production system (Stilla 1995). The interaction and interdependencies between objects in the store and productions is displayed by the production net (see Fig. 2).

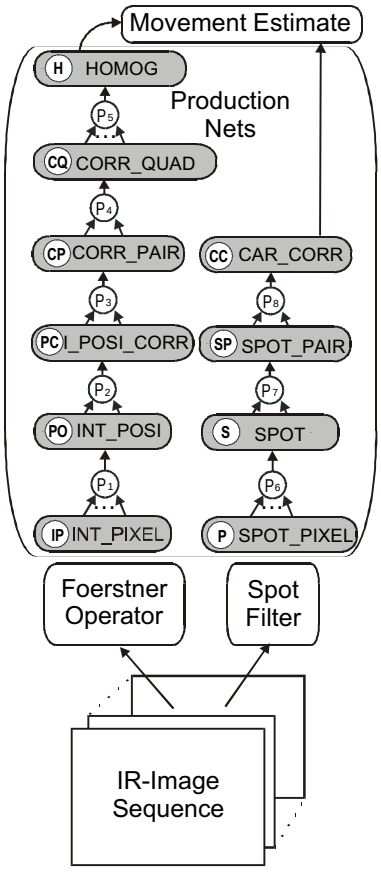

Figure 2. Production nets for GSAC-estimation of homographies and vehicle movement detection

Production p_1 combines clusters of objects INT_PIXEL of the image resulting from an operator introduced by Foerstner (1994). The resulting objects INT_POSI are located with sub-pixel accuracy due to averaging over the positions in a cluster weighted by the strength of the participating objects INT_PIXEL. The objects INT_POSI are assessed according to the overall mass of the underlying cluster. Production p_2 performs the search for corresponding objects INT_POSI from the other

frame of the image pair. If there is a suitable prior estimation of the homography (e.g. from the processing of preceding frames or external information), there may be quite narrow search regions to avoid excessive computation effort. The frames must not be directly adjacent in time. The resulting object INT_POSI_CORR is assessed using normalized grey-level correlation. These objects are used for estimations with IRLS and RANSAC, too.

Production p_3 and p_4 construct quadruple objects CORR_QUAD. Objects CORR_PAIR are used as intermediate step. They are assessed according to their distance, so that wide configurations covering as much of the image as possible are preferred. Objects CORR_QUAD are assessed using the area criterion q (Eq. 4). The estimated homography matrix resulting from their position correspondences is stored as attribute. Production p_5 operates on this matrix attribute domain and searches for a consensus set, i.e. a set of objects CORR_QUAD with similar matrix attributes. The matrices are stored with h_{33} scaled to unity to remove the homogenous ambiguity. The neighbourhood for the similarity relation is defined logarithmically. This is preferred due to the different sizes of the entries. The translation entries h_{13} and h_{23} can be fairly large. The projective entries h_{31} and h_{32} are usually quite small, but their sign is important.

The resulting object HOMOG is assessed according to the number and assessments of its predecessors. The estimation of the homography itself is done using squared error sum minimization over the set of correspondence objects preceding the quadruples that directly contributed to the cluster.

Productions p_6 to p_8 search for vehicle cues. The corresponding objects SPOT_PIXEL are based on a different pre-processing filter. Cars are formed from spot shaped objects which are extracted by a spot detector (own citation).

Productions p_6 and p_8 are similar to productions p_1 and p_2. Production p_7 assembles configurations of a cold and a hot spot close to each other into objects SPOT_PAIR. IR-images of urban areas contain many of these objects which do not result from vehicles. The evidence for objects CAR_CORR to represent a vehicle depends highly on its residual movement with respect to the best current estimation object HOMOG.

GSAC requires a data-driven control on the processing sequence. The application of productions is purely bottom-up. But, there is one exception. An object CAR_CORR that had a considerable residual movement earlier in the search may be rejected later, because an improved estimation object HOMOG reveals it as being stationary.

3.3 Exploitation of Context and Additional Knowledge

The combination of production systems with robust estimation of geometric entities allows including additional data, constraints, and knowledge into the estimation.

3.3.1 Digital Maps: Often there will be GIS-information available for the terrain which is observed. The building layer of such maps can be utilized to determine the correct setting in world coordinates. The road layer can be utilized to construct regions of interest for the vehicle search.

3.3.2 Data from Inertial Systems and the Flight Control: Often there will be data available about the aircrafts current position and heading, flying height and speed over ground, angular positions and changes in all three rotation axis. These data can be transformed into a priori homographies from one image to another and between images and GIS data. These can be used as expectations focussing the search on calculations that are probably less erroneous. Particularly, if there is evidence for no roll or nod rotations from the flight control, we may only accept small projective entries h_{31} and h_{32}.

4. RESULTS AND DISCUSSION

4.1 Comparing IRLS, RANSAC and GSAC

We applied different strategies to video sequences taken by a thermal camera from an aeroplane cruising over an urban area.

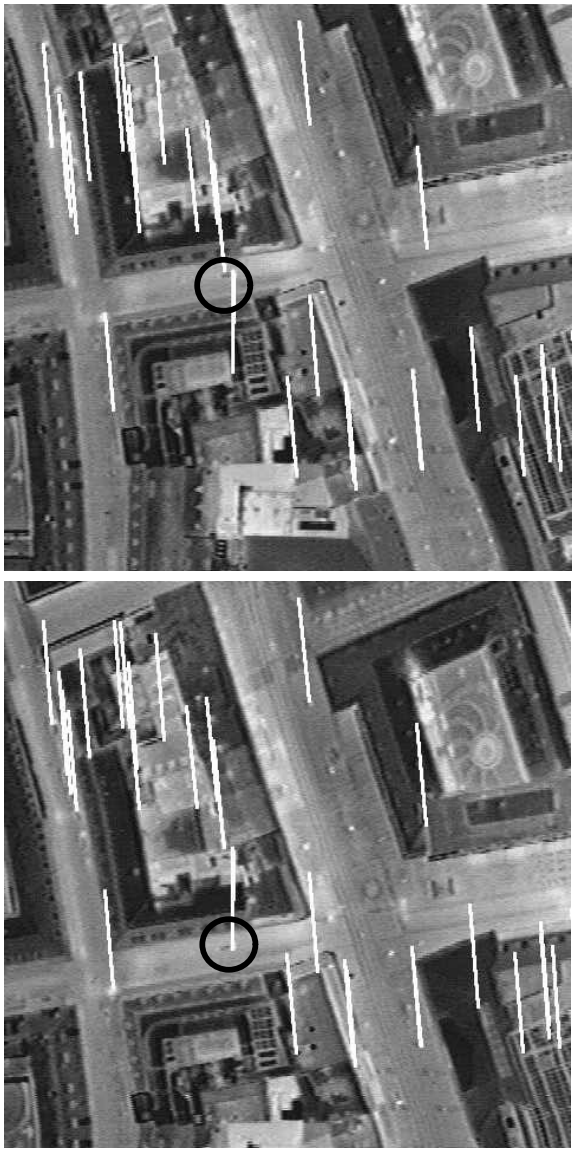

Figure 3. All correspondences in two example frames from a thermal image video (note the outlier at the moving car)

Fig. 3 displays a set of correspondence objects INT_POSI_CORR between two images with a time difference of 10 frames. The

slow moving vehicle on the horizontal road caused two outliers. Recall that correspondence of objects INT_POSI must not be unique. In this example we obtained two correspondences from one interest position. The vehicle is moving fast enough and the frames are far enough apart, so that these correspondences are easily detected as outliers by all three strategies. We assess the methods due to their choice of the correct correspondences they tend to reject.

4.1.1 IRLS: In Fig. 4 we displayed correspondences that have at least 1/7 of the weight of the best correspondence. The iterations were stopped after 20 steps. It shows that IRLS tended to put the highest weights on correspondences in the upper right triangular area.

One important correct correspondence on a bright spot (see Fig. 4, lower left area) gains a weight of less than 1/9 of best correspondence. The resulting homography-estimation yields highest precision in the middle of the region covered by the displayed sample. The moving car is located outside of this region. Still it is sufficient to detect the movement.

Figure 4. Correspondences preferred by IRLS

4.1.2 RANSAC: Vehicles may move arbitrarily slow. Therefore, a reasonable threshold for RANSAC-decisions needs statistic investigation on velocities of cars in the observed region. The behaviour of RANSAC highly depends on the choice of the threshold parameter.

For the run displayed in Fig. 5 we chose a threshold of 4 Pixel for the maximal displacement. 21 quadruples of correspondences were chosen at random. The one with the highest consensus is highlighted by surrounding the bright lines with a black margin. The white lines mark the consensus set. One important correspondence on the building roof vertex (Fig. 5, upper right area) is missing in this set (comp. Fig. 3 and Fig. 4). Recall that RANSAC is performing hard decisions. All correspondences in the set gain equal weight, while the others don't count at all. This makes such a rejection more serious than with the soft weighting in IRLS. For the purpose of estimating the velocity of the vehicle this set happens to be quite good, because the vehicle is within the convex hull of the positions of the correspondences.

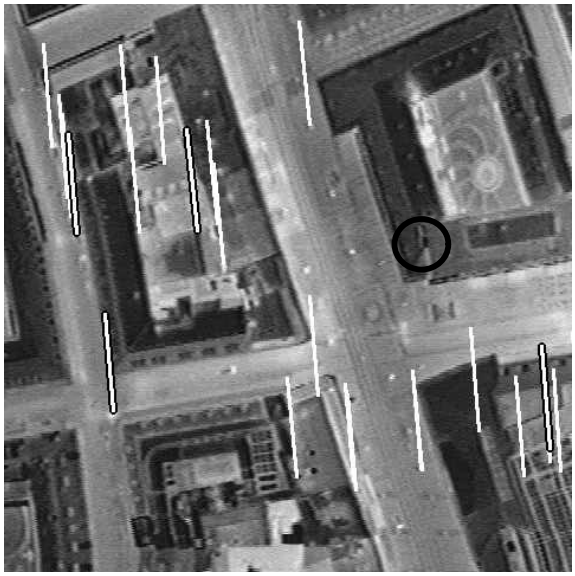

Figure 5. RANSAC consensus set with underlying sample emphasized

4.1.3 GSAC: Fig. 6 shows the correspondence objects INT_POSI_CORR preceding the homography estimation object HOMOG which has the best assessment.

Figure 6. GSAC consensus cluster

This sample only contains 12 correspondence objects INT_POSI_CORR, but these are spread all over the image. The homography estimation from this sample is very well suited for detecting moving vehicles.

4.2 Discussion

For robust estimation the standard deviation of the positioning of features has to be determined. The accuracy of the correlation supported location procedure should be below one

pixel. But there are other error sources, e.g. distortions from the camera system and effects from de-interlacing the video. From practical experience we assume a standard deviation $\sigma \approx$ 1.5 Pixel. Hartley & Zisserman (2000) recommend a threshold of 6σ for the RANSAC method. In this application such a threshold will sometimes collect outliers into the consensus set. Occasionally, experiments have shown such inclusions even with the threshold set to 6 Pixel. We admit, that in most cases with a threshold of 8 or 9 Pixel the outliers on the car where removed and the consensus contained all other correspondences.

A disadvantage of RANSAC for our task is the hard partitioning of the set of correspondences into two disjunct subsets, the inliers and the outliers. Many of the interest points result from 3D structures that will vary in appearance slightly with the view direction. Jet the scene has not enough depth to justify a 3D reconstruction e.g. using fundamental matrix estimation. Some of the interest points are very good for the estimation of 2D-homographies, especially if they are well spread all over the image. Some of them are still good enough to be considered. They improve the result. Some of them are surplus. Omitting them will not significantly change the result. And some of them look good at first glance but contradict the others (real outliers).

IRLS is a rather crude way to avoid the hard discrimination into inliers and outliers. The assessment of correspondences for the decision whether they should be included in the error minimization or should not be included should rather consider a combination of evidence. This includes evidence on the correspondence itself, quality or strength of the interest points on which it is based, and worth of it for the computation due to its position in the image. GSAC strategy is tailored to this combination of evidence.

Lately, there are several groups working on the improvement of RANSAC. A key issue is the assessment of samples according to their worth for the task at hand. In fact many implementations of RANSAC may already contain tacit preferences e.g., avoiding collinear or narrow samples. There are also publications treating this topic explicitly. Torr & Davidson (2003) adapt a method known from numerical integration as SIR (sampling – importance – re-sampling) to the task of fundamental matrix estimation using probability calculus.

Chum et al. (2003) provide a version of RANSAC which is `locally optimized´. A promising high scoring consensus sample is again re-searched for an `optimal´ minimal sub-sample. Particularly if the ratio of outliers to inliers is bad and if the error on the inliers is high, this will speed up the process dramatically. Besides, they recommend to use correspondences of image structures instead of simple points. Such a correspondence of what they call `distinguished regions´ provides also a local affine mapping. Thus only three correspondences form a minimal sample for fundamental matrix estimation.

For a proper selection of one of these different new assessment criteria and strategies for picking samples a comparison on common data sets from different tasks is necessary.

REFERENCES

Ben-Ezra, M., Peleg, S., Werman, M., 1999. Real time motion analysis with linear programming. *CVIU*, Vol.78, pp. 32-52.

Chum, O., Matas, J., Obdrzalek, S., 2003. Epipolar geometry from three correspondences. In: Drpohlav, O. (ed.): Computer vision – CVWW´03, Czech Pattern Recognition Society, Prague, pp. 83-88.

Dreschler, L., Nagel, H.-H., 1982. Volumetric model and trajectory of a moving car derived from monocular TV frame sequence of a street scene. *CGIP*, Vol. 20, pp. 199-228.

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.

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.

Haag, M., Nagel, H.-H., 1999. Combination of edge element and optical flow estimates for 3D-model based vehicle tracking in traffic image sequences. *IJCV*, Vol. 35:3, pp. 295-319.

Hartley, R., Zisserman, A., 2000. *Multiple view geometry*. 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., 2001. Estimating urban activity on high-resolution thermal image sequences aided by large-scale vector maps. In: *IEEE/ISPRS Joint Workshop URBAN´01*, pp. 25-29.

Partsinevelos, P., Agouris, P., Stefanidis A. 2000. Modelling movement relations in dynamic urban scenes. *International archives of photogrammetry and remote sensing*. Vol. 33, part B4, pp. 818-825.

Robust Estimation Library, ISBE, university of Manchester http://www.isbe.man.uk/public-vxl_doc/contrib/rpl/rrel/html/ (accessed 12 Mar. 2003)

Stilla, U., 1995. Map-aided structural analysis of aerial images. *ISPRS Journal of Photogrammetry and Remote Sensing*, Vol. 50, No. 4, pp. 3-10.

Torr, P. H. S., Davidson, C. 2003. IMPSAC: Synthesis of importance sampling and random sample consensus. *PAMI*, Vol. 25, No. 3, pp. 354-364.