Computer Vision — CVWW'03, Ondřej Drbohlav (ed.)
Valtice, Czech Republic, February 3–6, 2003
Czech Pattern Recognition Society

[←]

Epipolar Geometry from Three Correspondences

Ondřej Chum[1], JiříMatas[1,2], Štěpán Obdržálek[1]

[1]Center for Machine Perception, Czech Technical University, Prague, 120 35, CZ
[2]Centre for Vision Speech and Signal Processing, University of Surrey, Guildford, GU2 7XH, UK
chum@cmp.felk.cvut.cz, matas@cmp.felk.cvut.cz, xobdrzal@fel.cvut.cz

Abstract *In this paper,* LO-RANSAC *3-LAF – a new algorithm for the correspondence problem – is described. Exploiting processes proposed for computation of affine-invariant local frames, three point-to-point correspondences are found for each region-to-region correspondence. Consequently, it is sufficient to select only triplets of region correspondences in the hypothesis stage of epipolar geometry estimation by* RANSAC.

We experimentally show that: 1. LO-RANSAC *3-LAF estimates epipolar geometry in time that is orders of magnitude faster than the standard method, 2. that the precision of the* LO-RANSAC *3-LAF and the standard method are comparable, and 3. that* RANSAC *without local optimisation applied to triplets of points from a single region is significantly less precise than the new* LO-RANSAC *3-LAF algorithm.*

In the experiments, a speed-up factor in orders of thousands is achieved on the problem of epipolar geometry estimation. The proposed method is pushing the limit of solvable problems, allowing EG estimation in correspondence problems with the number of inliers below 10%.

1 Introduction

Establishing correspondence in a pair of images is an important component of many computer vision systems. In this paper we focus on the two-view (stereo) problem, however all the results have direct impact on the multi-view and object recognition problems.

Usually, especially in the narrow-baseline stereo, epipolar geometry (EG) has been estimated from sets of point-to-point correspondences. This is a consequence of two factors: 1. The almost universal use of Harris point detector, and 2. the good understanding of the estimation of fundamental matrix from point correspondences. Interest point detectors, such as Harris, commonly assume approximations of local image deformations as translation and rotation. Their output depends on properties of fixed-sized circular regions to achieve the invariance to the image transformations. In a general correspondence problem, it is necessary to model local image transformation as affine. Consequently, detectors of entities that are put into correspondence must be more complex. The output of these detectors is a distinguished *region* of data dependent shape, detected in an affine-invariant manner. Such detectors have been proposed by Tuytelaars [13], Prittchet [8], Schaffalitzky [9] and Matas [5].

A globally consistent set of correspondences of the distinguished regions is found essentially in two steps. First, tentative correspondences are selected by comparing locally computed (invariant) descriptors characterising the distinguished regions. In the second step, a maximal mutually consistent subset of tentative correspondences is found, typically by RANSAC. RANSAC proceeds in a hypothesise-and-verify manner as follows: 1. the smallest sample sufficient to instantiate a model is selected randomly from the set of tentative correspondences, 2. a model (e.g. the fundamental matrix) is computed from the correspondences, 3. the model is verified, i.e. the number of tentative correspondences consistent with the hypothesised model is computed.

When can a correspondence problem be solved? Firstly, in the detection phase, a sufficient *absolute* number of detected regions in both images must correspond to the same physical surface patches. Secondly, the locally computed description associated with a region must be discriminative enough to ensure that a sufficient *proportion* ε of tentative correspondences is correct. The RANSAC time complexity is a high degree polynomial of ε; the degree of the polynomial depends on the minimum sample size needed to instantiate a model. Small ε results in enormous growth of the number of hypotheses that will be evaluated.

So far, all published work on wide-baseline stereo has used a single point-to-point geometric constraint per tentative correspondence. This means that seven correspondences have been needed to estimate the fundamental matrix and the expected number of tested hypotheses is proportional to $1/\varepsilon^7$ [4]. A *single* point-to-point correspondence has been used, despite the fact that stronger geometric constrains can be derived from a region-to-region correspondence. In fact, in the work of Tyutelaars [13] and Pritchett [8], parameters of local region-to-region mapping have been estimated and used in the process of selection of tentative correspondences. However, epipolar geometry estimation ignored this information and the fundamental matrix has been computed from seven point correspondences. A possible reason is that the fundamental matrix estimated from small number of local measurements is not as precise as the one obtained from seven independent (and thus spatially randomly distributed) correspondences.

*The authors were supported by the European Union under project IST-2001-32184, by the Czech Ministry of Education under project MSM 212300013 and by The Grant Agency of the Czech Republic under projects GACR 102/02/1539

In a main contribution of this paper, a new algorithm for the correspondence problem is proposed. Exploiting processes proposed for computation of affine-invariant local frames [6], three point-to-point correspondences are found for each region-to-region correspondence. It is therefore sufficient to select only a triplet of region correspondences in the hypothesis stage of the RANSAC and the expected runtime falls to $t \approx \frac{1}{\varepsilon^m}$, where $m = 3$.

Since the RANSAC running time depends on the minimal sample size m (the 'model dimensionality') *exponentially*, the effect of this modification leads to speed-up of several orders of magnitude. To appreciate the effect, let us consider $\varepsilon = 0.15$, i.e. 15% of inliers. The proposed algorithm has $m = 3$ as opposed to the standard $m = 7$. The respective times are $t_{m=3} = 297$ for the proposed algorithm and $t_{m=7} = 585277$ for the standard RANSAC. The straightforward consequence is an enormous enlargement of the class of problems that are solvable.

The idea of using multiple points in the estimation process is in principle simple. However, since the three points associated with a single region are in close proximity, the precision of the estimated epipolar geometry may be questioned. Indeed, we have observed experimentally, that the estimated fundamental matrix is imprecise to a point where a significant proportion of correct tentative correspondences is inconsistent with the estimated model. The new approach seems to have two problems. Firstly, the precision is traded off for speed. Secondly, since the termination of RANSAC depends on the proportion of tentative correspondences consistent with the hypothesised model, the theoretical speed-up is not achieved.

Both these problems are alleviated by replacing the RANSAC with the locally optimized RANSAC (LO-RANSAC) recently introduced by Chum and Matas [1]. LO-RANSAC solves both problems by maximizing the number of inliers consistent with the hypothesized model by local optimisation. The optimisation causes only negligible slowdown in overall performance, as it is carried out infrequently (see Equation 3 and [1]).

The outline of the matching process is as follows:

1. For both images compute distinguished regions, establish local affine frames, and generate intensity representation of local image patches normalised according to the local frames.
2. Establish tentative correspondences between the frames, by directly comparing the normalised local image intensities.
3. Select a maximal mutually consistent set of tentative correspondences by applying the locally optimised RANSAC, the LO-RANSAC, to triplets of frame-to-frame correspondences.

The rest of the paper is structured as follows: the concept of *distinguished regions* is briefly explained in Section 2. The description of local affine frames follows in Section 3 and the details on the LO-RANSAC algorithm are given in Section 4. Experiments on wide-baseline image pairs are shown in Section 5, and the paper is concluded and the contributions are summarized in Section 6.

2 Distinguished Regions

Distinguished Regions (DRs) are image elements (subsets of image pixels), that posses some distinguishing, singular property that allows their repeated and stable detection over a range of image formation conditions. In this work we exploit a new type of distinguished regions introduced in [5], the *Maximally Stable Extremal Regions* (MSERs). An extremal region is a connected component of pixels which are all brighter (MSER+) or darker (MSER-) than all the pixels on the region's boundary. This type of distinguished regions has a number of attractive properties: 1. invariance to affine and perspective transforms, 2. invariance to monotonic transformation of image intensity, 3. computational complexity almost linear in the number of pixels and consequently near real-time run time, and 4. since no smoothing is involved, both very fine and coarse image structures are detected. We do not describe the MSERs here; the reader is referred to [5] which includes a formal definition of the MSERs and a detailed description of the extraction algorithm. Examples of detected MSERs are shown in Figure 1. Note that DRs do not form segmentation, since DRs do not cover entire image area, and DRs can be (and usually are) nested.

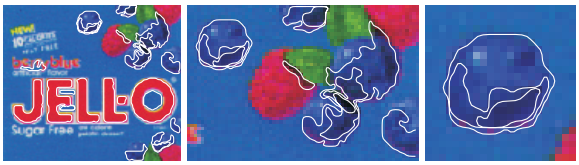


Figure 1: An example of detected regions of MSER type

3 Local Frames of Reference

Local affine frames (LAFs) facilitate normalisation of image patches into a canonical frame and enable direct comparison of photometrically normalised intensity values, eliminating the need for invariants. It might not be possible to construct local affine frames for every distinguished region. Indeed, no dominant direction is defined for elliptical regions, since they may be viewed as affine transformations of circles, which are completely isotropic. On the other hand, for some distinguished regions of a complex shape, multiple local frames can be affine-covariantly constructed in a stable and thus repeatable way. Robustness of our approach is thus achieved by selecting only stable frames and employing multiple processes for frame computation.

Definition of terms:

Affine transformation is a map $F : \mathbb{R}^n \to \mathbb{R}^n$ of the form $F(\mathbf{x}) = A^T\mathbf{x} + \mathbf{t}$, for all $\mathbf{x} \in \mathbb{R}^n$, where A is a linear transformation of $\mathbb{R}^n$, assumed non-singular here.

Center of gravity (CG) of a region Ω is $\mu = \frac{1}{|\Omega|}\int_\Omega \mathbf{x} d\Omega$.

Covariance matrix of a region Ω is a $n \times n$ matrix defined

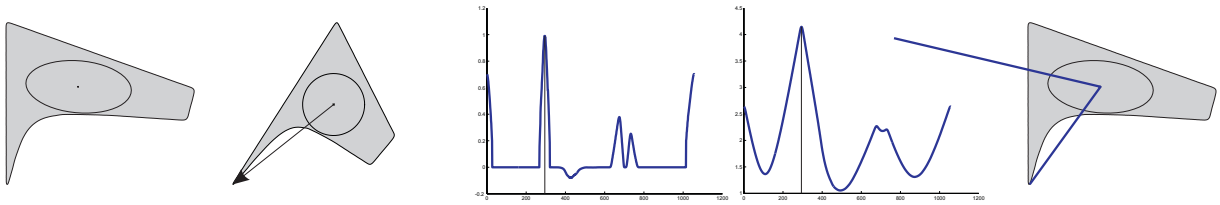
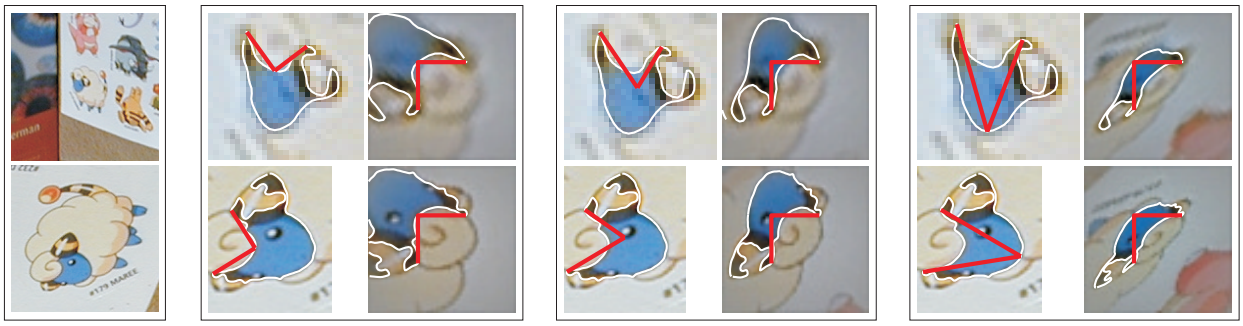


Figure 2: Construction of affine frames. From left to right: a distinguished region (the gray area), the DR shape-normalised according to the covariance matrix, normalised contour curvatures, normalised contour distances to the center of DR, and one of the constructed frames represented by its basis vectors.



(a) (b) (c) (d)

Figure 3: Bi-tangent based constructions of affine frames. (a) original views, (b) 2 tangent points + farthest concavity point, (c) 2 tangent points + DR's center of gravity, (d) 2 tangent points + farthest DR point. Left columns - detected frames, right columns - locally normalised images

as
$\Sigma = \frac{1}{|\Omega|} \int_\Omega (\mathbf{x} - \mu)(\mathbf{x} - \mu)^T \, d\Omega$.

Bi-tangent is a line segment bridging a concavity, i.e. its endpoints are both on the region's outer boundary and the convex hull, all other points are part of the convex hull.

Affine covariance of the center of gravity and of the covariance matrix is shown in [6]. The invariance of the bi-tangents is a consequence of the affine invariance (and even projective invariance) of the convex hull construction [10, 7]. Finally, we exploit the affine invariance of the maximal-distance-from-a-line property, which is easily appreciated taking into account that affine transform maintains parallelism of lines and their ordering.

A two-dimensional affine transformation possesses six degrees of freedom. Thus, to determine an affine transformation, six independent constraints are to be applied. Various constructions can be utilised to obtain these constraints. In particular, we use a direction (providing a single constraint), a 2D position (providing two constraints), and a covariance matrix of a 2D shape (providing three constraints).

Frame constructions. Two main groups of affine-invariant constructions are proposed, based on 1. region normalisation by the covariance matrix and the center of gravity, and 2. detection of stable bi-tangents

Transformation by the square root of inverse of the covariance matrix normalises the DR up to an unknown rotation. To complete an affine frame, a direction is needed to resolve the rotation ambiguity. The following directions are used: 1. Center of gravity (CG) to a contour point of extremal (either minimal or maximal) distance from the CG 2. CG to a contour point of maximal convex or concave curvature, 3. CG of the region to CG of a concavity, 4. direction of a bi-tangent of a region's concavity.

In frame constructions derived from the bi-tangents, the two tangent points are combined with a third point to complete an affine frame. As the third point, either 1. the center of gravity of the distinguished region, 2. the center of gravity of the concavity, 3. the point of the distinguished region most distant from the bi-tangent, or 4. the point of the concavity most distant from the bi-tangent is used. Another type of frame construction is obtained by combining covariance matrix of a concavity, CG of the concavity and the bi-tangent's direction.

Frame constructions involving the center of gravity or the covariance matrix of a DR rely on the correct detection of the DR in its entirety, while constructions based solely on properties of the concavities depend only on a correct detection of the part of the DR containing the concavity.

Figure 2 visualise the process of shape-normalisation and a dominant point selection. A distinguished region detected in an image is transformed to the shape-normalised frame, the transformation being given by the square root of inverse of the covariance matrix. Normalised contour curvatures and normalised contour distances are searched for stable extremal values to resolve the rotation ambiguity. One of the constructed frames is shown on the right in Figure 2, represented by the two basis vectors of the local coordinate system. Figure 3 shows three examples of the local affine frame constructions based on concavities.

Once local affine frames (LAFs) are computed in a pair of images, (geometrically) invariant descriptors of local appearance are not needed for the matching. Correspondences

are established simply by correlating photometrically normalised image intensities in geometrically normalised measurement regions. A measurement region (MR) is defined in local coordinate systems of the affine frames, but the choice about MR shape and size can be arbitrary. Larger MRs have higher discriminative potential, but are more likely to cover an object area that violates the local planarity assumption. Our choice is to use a square MR centered around a detected LAF, specifically an image area spanning $[-2, 3] \times [-2, 3]$ in the frame coordinate system.

4 Locally Optimized RANSAC

The structure of the RANSAC algorithm is simple but powerful. Repeatedly, subsets are randomly selected from the input data and model parameters fitting the sample are computed. The size of the random samples is the smallest sufficient for determining model parameters. In a second step, the quality of the model parameters is evaluated on the full data set. The process is terminated [2, 11] when the likelihood of finding a better model becomes low, i.e. the probability η of missing a set of inliers of size I within k samples falls under a predefined threshold

$$\eta = (1 - P_I)^k. \tag{1}$$

Symbol P_I stands for the probability, that an uncontaminated sample of size m is randomly selected from N data points

$$P_I = \frac{\binom{I}{m}}{\binom{N}{m}} = \prod_{j=0}^{m-1} \frac{I-j}{N-j} \approx \varepsilon^m, \tag{2}$$

where ε is the fraction of inliers $\varepsilon = I/N$. The number of samples that has to be drawn to ensure given η is

$$k = \log(\eta)/\log(1 - P_I).$$

From equations (1) and (2), it can be seen, that termination criterion based on probability η expects that a selection of a single random sample not contaminated by outliers is followed by a discovery of whole set of I inliers. However, this assumption is often not valid since inliers are perturbed by noise. Since RANSAC generates hypotheses from minimal sets, the influence of noise is not negligible, and a support set of correspondences with size smaller than I is found. The consequence is an increase in the number of iterations before the algorithm is terminated.

We propose a RANSAC modification that increases the number of inliers found near to the optimum I. This is achieved by local optimisation of 'promising' samples. For the overview of the locally optimized RANSAC see Algorithm 1.

The local optimization step is carried out only if a new maximum in the number of inliers from the current sample has occurred, i.e. when standard RANSAC stores its best result. The number of consistent data points with a model from a randomly selected sample can be thought of as a random variable with unknown (or very complicated) density function. This density function is the same for all samples, so the probability that k-th sample will be the best so far is

Repeat until the probability of finding better solution falls under predefined threshold, as in equation (1):

1. Select a random sample of the minimum number of data points S_m.
2. Estimate the model parameters consistent with this minimal set.
3. Calculate the number of inliers I_k, i.e. the data points their error is smaller than predefined threshold θ.
4. If new maximum has occurred ($I_k > I_j$ for all $j < k$), run **local optimization**. Store the best model.

Algorithm 1: A brief summary of the LO-RANSAC

$1/k$. Then, the average number of times a new maximum is found within k samples is

$$\sum_{1}^{k} \frac{1}{x} \leq \int_{1}^{k} \frac{1}{x}\, dx + 1 = \log k + 1. \tag{3}$$

Model hypothesizing. To hypothesize a model of epipolar geometry, random samples of three region correspondences are drawn. Each region correspondence provides a point and a local affine transformation. Generally, three region correspondences give nine point correspondences. These are used to estimate the fundamental matrix F using linear eight-point algorithm [3].

Verification and local optimization. Both verification and local optimization are run on point correspondences only. This is due to higher stability of the central point as the aim of the local optimization is higher precision of the fundamental matrix.

The set of inliers is calculated from the hypothesized fundamental matrix. If a new maximum of inliers is reached, the local optimization is carried out as follows. A new sampling procedure is run only on I_k data points consistent with the hypothesized model. As the sampling is running on inlier data, there is no need for the size of the sample to be minimal. On the contrary, the size of the sample is selected to minimize the error of the model parameter estimation. In our experiments the size of samples are set to $\min(I_k/2, 14)$ The number of repetitions is set up to twenty in the experiments presented. Each random sample drawn in this procedure is processed with the following iterative scheme: take all data points with error smaller that $K \cdot \theta$ and use linear algorithm to compute new model parameters. Reduce the threshold and iterate until the threshold is θ. This is equivalent to the most efficient method described in [1].

Rejecting random inliers. The affine transformation can be used to reject random inliers, i.e. regions that 'randomly' lay on corresponding epipolar lines. For this purpose the error of epipolar geometry on the other two points (defining the local affine transformation of the region correspondence) can be thresholded, typically by a weaker threshold than θ.

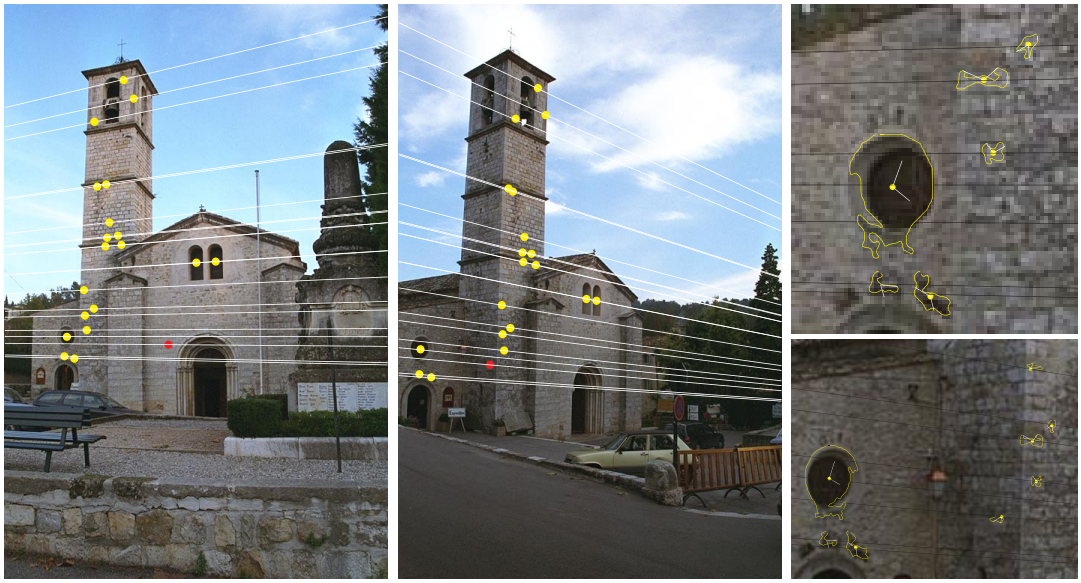


Figure 5: Epipolar geometry estimated on the 'Valbonne' image pair. A detail with frame and region correspondences is shown on the right.

Figure 4: Epipolar geometry estimated on the 'Wash' image pair.

Method	EG consistent	iterations
Wash, **191** tentative corr., **42**% inliers		
RANSAC 7-pts	80	2178
RANSAC 3-LAF	47	475
LO-RANSAC 3-LAF	80	35
Valbonne, **193** tentative corr., **12**% inliers		
RANSAC 7-pts	22	≈ 20 000 000
RANSAC 3-LAF	15	8608
LO-RANSAC 3-LAF	22	869
Ascona, **783** tentative corr., **13**% inliers		
RANSAC 7-pts	100	≈ 6 500 000
RANSAC 3-LAF	33	67325
LO-RANSAC 3-LAF	102	1105

Table 1: Summary of experimental results. Number of correspondences found consistent with the epipolar geometry and the number of RANSAC iterations required to reach the solution. Note that all the numbers are random variables.

5 Experiments

In this section we experimentally demonstrate the benefits of the newly proposed LO-RANSAC 3-LAF method. In particular we show that: 1. LO-RANSAC 3-LAF estimates epipolar geometry in time that is orders of magnitude faster than the standard method, i.e. RANSAC 7-pts, 2. that the precision of the LO-RANSAC 3-LAF and RANSAC 7-pts are comparable, and 3. that RANSAC without local optimisation applied to triplets of points from a single region, RANSAC 3-LAF, is significantly less precise than both LO-RANSAC 3-LAF and RANSAC 7-pts.

To highlight the advantage of the proposed approach, two complicated image pairs are included with only about 12% of tentative correspondences correct. The image pairs are depicted in Figures 5 and 6 respectively. A simple experiment on an indoor scene (Figure 4) represents a fairly standard wide-baseline problem. Information about the conducted experiments is summarised in Table 1. For the 'Valbonne' and 'Ascona' pairs, the speed-up measured by the number iterations is formidable, see the left column of Table 1. Due to the low probability of picking seven correct correspondences in the conventional seven-point algorithm (RANSAC 7-pts), the number of iterations is in the order of tens of millions. In comparison, only about a thousand of iterations is required by the proposed LO-RANSAC 3-LAF method. For the 'Wash' experiment, the number of iterations fall down to 35.

The quality of the EG estimate measured by the number of inliers (see the middle column of Table 1) is comparable for the standard RANSAC 7-pts and the proposed LO-RANSAC 3-LAF methods. Figures 4, 5 and 6 show the epipolar geometry superimposed over the images. The quality of the geometry can be appreciated by looking at the close-ups in Figures 4 and 5. The close-ups also show the distinguished regions and the basis vectors of the corresponding local affine frames.

Figure 6: Epipolar geometry estimated on the 'Ascona' image pair.

It is important to note that applying the 3-frame RANSAC without local optimisations (RANSAC 3-LAF), the largest set of EG-consistent inliers is smaller. We believe that this is due to the fact that local affine frames are typically very small and the three points from a single region lie in close proximity. The local inter-image affine transformation are thus inaccurate and consequently the EG estimate and its support set (the set of inliers consistent with the EG estimate) are unstable. The size of support sets for RANSAC 3-LAF and LO-RANSAC 3-LAF can differ significantly, as shown in the middle column of Table 1.

The difference in the support set size also affects the speed of the algorithm, since the termination condition of RANSAC is a function of the current estimate of the number of inliers. As a consequence of its lower precision (smaller support set), RANSAC 3-LAF was more than ten times slower than LO-RANSAC 3-LAF in all conducted experiments (Table 1, right column). Clearly, application of local optimisation is a very important ingredient of the newly proposed algorithm.

In the experiments, *Maximally Stable Extremal Regions* (MSERs) [5] were used as distinguished regions. All the affine-covariant constructions described in Section 3 were exploited.

6 Conclusions

In this paper, LO-RANSAC 3-LAF – a new algorithm for the correspondence problem – was described. Exploiting processes proposed in [6] for computation of affine-invariant local frames, three point-to-point correspondences were found for each region-to-region correspondence. Consequently, it was sufficient to select only triplets of region correspondences in the hypothesis stage of epipolar geometry estimation by RANSAC. In the experiments, a speed-up factor in order of thousands was achieved on the problem of epipolar geometry estimation.

The importance of local optimisation in the RANSAC verification stage for both precision and speed of the method was demonstrated. Local optimisation helped to overcome the problem of precision of estimation from points in close proximity. The computed epipolar geometry was as precise as if using the conventional algorithm which exploits 7 points in general positions.

The proposed method is pushing the limit of solvable problems, allowing EG estimation in correspondence problems with the ratio of inliers below 10%.

References

[1] O. Chum, J. Matas, and J. Kittler. Locally optimized RANSAC: Running as standard RANSAC should?, submited to CVPR'03, 2003.

[2] M. Fischler and R. Bolles. Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography. *CACM*, 24(6):381–395, June 1981.

[3] R. Hartley. In defence of the 8-point algorithm. In *ICCV95*, pages 1064–1070, 1995.

[4] R. Hartley and A. Zisserman. *Multiple View Geometry in Computer Vision*. Cambridge University Press, Cambridge, UK, 2000.

[5] J. Matas, O. Chum, M. Urban, and T. Pajdla. Robust wide baseline stereo from maximally stable extremal regions. In *Proceedings of the British Machine Vision Conference*, volume 1, pages 384–393, 2002.

[6] J. Matas, Štěpán Obdržálek, and O. Chum. Local affine frames for wide-baseline stereo. In *ICPR02*, August 2002.

[7] J. L. Mundy and A. Zisserman, editors. *Geometric Invariance in Computer Vision*. The MIT Press, 1992.

[8] P. Pritchett and A. Zisserman. Wide baseline stereo matching. In *Proc. International Conference on Computer Vision*, pages 754–760, 1998.

[9] F. Schaffalitzky and A. Zisserman. Viewpoint invariant texture matching and wide baseline stereo. In *Proc. 8th International Conference on Computer Vision, Vancouver, Canada*, July 2001.

[10] T. Suk and J. Flusser. Convex layers: A new tool for recognition of projectively deformed point sets. In F. Solina and A. Leonardis, editors, *Computer Analysis of Images and Patterns : 8th International Conference CAIP'99*, number 1689 in Lecture Notes in Computer Science, pages 454–461, Berlin, Germany, September 1999. Springer.

[11] P. Torr, A. Zisserman, and S. Maybank. Robust detection of degenerate configurations while estimating the fundamental matrix. *CVIU*, 71(3):312–333, September 1998.

[12] P. H. S. Torr and A. Zisserman. MLESAC: A new robust estimator with application to estimating image geometry. *Computer Vision and Image Understanding*, 78:138–156, 2000.

[13] T. Tuytelaars and L. Van Gool. Wide baseline stereo matching based on local, affinely invariant regions. In *Proc. 11th British Machine Vision Conference*, 2000.