ČESKÉ VYSOKÉ UČENÍ TECHNICKÉ
V PRAZE

TEZE K DISERTAČNÍ PRÁCI

České vysoké učení technické v Praze
Fakulta elektrotechnická
Katedra kybernetiky

Michal Perďoch

Maximally Stable Extremal Regions and Local Geometry for Visual Correspondences

Doktorský studijní program: Elektrotechnika a informatika, P2612

Studijní obor: Umělá inteligence a biokybernetika, 3902V035

Study Programme No. P 2612 - Electrotechnics and Informatics, branch No. 3902V035 - Artificial Intelligence and Biocybernetics

Teze disertace k získání akademického titulu "doktor", ve zkratce "Ph.D."

Praha, srpen 2011

Disertační práce byla vypracována v rámci prezenční formy doktorského studia na pracovišti Centra strojového vnímání na katedře kybernetiky Fakulty elektrotechnické ČVUT v Praze.

Disertant: Ing. Michal Perďoch
Katedra kybernetiky FEL ČVUT

Školitel: Prof. Dr. Ing. Jiří Matas
Katedra kybernetiky FEL ČVUT

Oponenti: ....

....

....

Teze byly rozeslány dne ....

Obhajoba se koná dne .... v .... hodin v zasedací místnosti číslo .... Fakulty elektrotechnické ČVUT v Praze před komisí pro obhajobu ve studijním oboru 3902V035, Umělá inteligence a biokybernetika.

S disertací je možno se seznámit na děkanátě elektrotechnické fakulty ČVUT v Praze, Technická 2, Praha 6 – Dejvice.

Prof. Ing. Vladimír Mařík, DrSc.
předseda komise pro obhajobu DP ve studijním
oboru 3902V035, Umělá inteligence a biokybernetika,
Katedra kybernetiky, Karlovo náměstí 13, Praha 2

Contents

1 Problem Formulation

The problem of finding correspondences between different views of an object or scene has been well studied in computer vision. Many algorithms need to establish correspondences of objects in the scene, including real-time tracking, wide baseline stereo, multiview geometry estimation, object detection and recognition, and large-scale image retrieval. Recent state-of-the-art methods [21, 12, 24, 26] for object recognition and image retrieval are based on establishing correspondences by matching invariant descriptions of local features. Objects or images are decomposed into smaller salient points or regions, called *local features*. The local features are represented by invariant descriptors and are matched on the basis of descriptor similarity. Subsequently, the matches are typically verified using a global geometry constraint.

The visual correspondence problem *formulated* as a search for a geometrically consistent set of matching descriptors of local features between the query and database images has following components: feature detector, feature description, matching or indexing, and verification. All of the components are already available for some classes of objects. However, they have limitations, some of which are addressed in this thesis. The *goals of the thesis* are to:

- Analyse the Maximally Stable Extremal Regions (MSERs) detector [14], propose improvements addressing its weaknesses and improve the computational costs.
- Propose a novel concept for affine covariant detection of stable parts of the extremal region boundary.
- Demonstrate the benefits of using local geometry constraints provided by affine covariant local features in efficient visual correspondence verification and symmetry detection scenarios.
- Develop an efficient representation of the local geometry constraints of affine covariant local features to meet the requirements of recent large-scale retrieval systems.

2 Contributions

The main contributions of the thesis are:

- Generalisations of the MSERs *stability function* used in the MSER detector are proposed. The performance is evaluated in a large-scale image retrieval application, and the repeatability measured using a standard protocol.

- The union-find based implementation of the MSER detector was analysed, critical performance parameters identified, and a set of improvements that *reduces the number of allocations and size of memory footprint* was proposed. The obtained implementation is showing benefits of the reduced memory footprint, optimisation of internal structures and improved output performance and achieves a two fold reduction of the execution time with respect to the baseline implementation [14].

- A novel concept of Stable Affine Frames is proposed that relaxes the requirement of a stable extremal region and enables the extraction of the stable part of the boundary of an extremal region.

- The visual correspondences of affine covariant features are exploited in a new approach for efficient estimation and verification of epipolar geometry. The approach uses RANSAC with local optimisation to robustly estimate the model of epipolar geometry from a *single pair* of affine correspondences. This allows lower order dependence on the inlier ratio in estimation of the epipolar geometry for the wide baseline stereo, object recognition and image retrieval.

- The estimation of bilateral symmetry from local features by Cornelius and Loy is extended to form the hypothesis of a bilateral symmetry from a single correspondence of affine covariant features. This reduces the number of possible hypothesis, and also speeds up the verification phase.

- Finally, a new method for representation of local geometry is proposed and its performance shown in a large-scale object retrieval application. The method enables storage of the local geometry in less than 24 bits per feature without a significant influence on the retrieval performance. Geometry and appearance information is achieved with less than 6 bytes of storage per feature enabling real-time retrieval in large ($> 10^6$) collections of images.

3 The State of the Art

This thesis addresses many aspects of the visual correspondence problem, some of the relevant references to each of them are listed in the following.

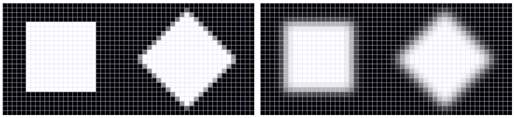
The Maximally Stable Extremal Regions (MSERs) has been used in a wide variety of computer vision tasks, including wide-baseline stereo [14], image retrieval [24, 18]), object recognition [20], and tracking and 3D segmentation [7]. The MSER detector has been implemented on FPGA as well. Several extensions of the MSERs were proposed, that enable processing of colour images [8], extend MSERs to a temporal dimension [7] and volume [6] or enable detection of class specific region [15]. Demand for processing of large-scale image databases (with more than 10^6 images) encourages research of computational costs [17, 19] of the baseline algorithm.

The boundaries of MSERs contain all parts of discrete intensity level sets, also denoted as discrete isophotes. Caselles *et al*. [1] studied the properties of the intensity level sets and shown that they contain enough information for reconstruction of the image. The stability of MSERs is influenced by the stability of the whole connected component, selection of repetitive affine covariant parts or features on the regions boundary was proposed by Lamdan *et al*. [11] and Obdrzalek *et al*. [20]. Both assume that whole boundary is available beforehand.

The seven point algorithm embedded in RANSAC is a standard method for epipolar geometry estimation in wide baseline stereo [14]. The speed of RANSAC is inversely proportional to an exponential function of the sample size. Chum and Matas [2, 3] studied whether the required seven point-to-point correspondences obtained from three correspondences of affine frames allows efficient estimation of epipolar geometry. Experiments in [3] show that standard RANSAC fails in this case. However, a simple modification called *local optimisation of the so-far-the-best solution* leads to an algorithm that benefits from the small sample size without losing efficiency; speed-ups of up to 10^3 were reported.

Symmetry detection has found use in numerous applications ranging from facial image analysis and vehicle detection to 3D reconstruction and visual attention. The local-feature based approaches offer numerous advantages. In particular, the ability to efficiently detect local symmetries in images that are not globally symmetric. Tuytelaars *et al*. [25] presented a method for the detection of regular repetitions of planar patterns under perspective skew using a geometric framework. Recently Cornelius and Loy [5] proposed a method for grouping symmetric constellations of features and detecting symmetry in perspective.

Geometric verification has been shown essential in recent state-of-the-art image retrieval approaches [4, 22, 9]. Local geometry represented by an affine covariant ellipse is noticeably bigger than the size of tf-idf weights and labels when stored in a naive way thus becoming a significant factor determining the



(a) (b)

Figure 1: Effects of finite resolution and blur on a square region. Left original image. Right image blurred by Gaussian blur. Mixed pixels (between black background and white object appeared).

limits of a method.

4 MSER Improvements

The *extremal regions* are defined as a connected components of pixels that are all of a either higher (maximal region) or lower (minimal region) value than the pixels on the boundary of the region. In other words, the extremal regions are sub-level sets of the intensity function (minimal regions) or sub-level sets of the reversed intensity function (maximal regions). The standard *maximally stable* extremal regions are selected among extremal regions by the *stability function*: $q(i) = |\mathcal{Q}_{i+\Delta} \setminus \mathcal{Q}_{i-\Delta}|/|\mathcal{Q}_i|$, where $\mathcal{Q}_i$ is extremal region at intensity i, $|.|$ denotes number of pixels. Extremal region $\mathcal{Q}_{i^*}$ is *maximally stable* iff has a local minimum at i^*. Extremal regions posses interesting properties, they are invariant to affine transformations of image intensities and covariant to any adjacency preserving continuous transformation of the image domain. Since no smoothing is involved, MSER detector concurrently finds both very fine and very large structures. MSERs can be enumerated in linear [18] or quasi-linear time [14] in the number of pixels. However, they also have few weaknesses [16] that include sensitivity to blur, discretisation effects or sparseness (non-extremality of interesting structures).

Discretisation effects. The change of region area is given by the number of pixels $\mathcal{Q}_{i+\Delta} \setminus \mathcal{Q}_{i-\Delta}$ denoted as *mixed pixels*. An examples are shown in Figure 1(a) and 1(b), the mixed pixels are the grey pixels on boundary of a white maximal region. Notice that even though the stability definition is affine invariant, the minimum value of $q(i)$ varies for stable regions of different sizes due to discretisation effects. Small regions tend to have a much higher relative number of mixed pixels on their boundary than big regions. In practise, one do not want to extract all local minima of $q(i)$ and picks the "stable" ones by setting a threshold on $q(i)$. However, the threshold on the value of $q(i)$ results in an underestimation of the stability due to discretisation effects on small regions.

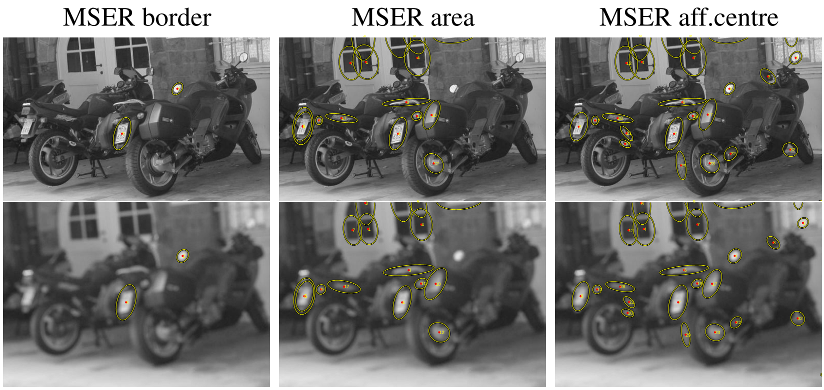


Figure 2: Analysis of the stability functions (cols) behaviour on blurred scene. Repeated features are for same values of the stability threshold. Note the absence of nested regions for the MSER aff. centre method.

Sensitivity to blur. Depending on the size of the square, resolution of the camera, and amount of blur, different a number of mixed pixels appear on the boundary of the region. Now let us assume some fixed threshold t on the value of $q(i)$. On an in-focus image of the scene (*c.f*. Figure 1(a)), the number of mixed pixels is proportional to the length of the boundary, and easily falls under the threshold t. In contrast to a blurred image (*c.f*. Figure 1(b), the number of mixed pixels grows significantly *e.g*. threefold which effectively increases $q(i)$ three times.

Stability based on the centroid of the region Suppose that the number of mixed pixels on the region boundary is function of the amount of blur. The amount of image blur cannot be easily determined, this would require to solve the blind deconvolution problem, although estimates from image gradients or other statistics may be computed. We are looking for a criterion of stability that exploits some property common to both stable and blurred regions. The stability of the centroid of the region, because the position of the centroids of the stable regions in the in-focus image will not differ much in the defocused image. Affine invariance can be added to the stability function by using the region's covariance matrix to measure the movement of the centroid position in the normalised coordinates. Several weaknesses of MSER detector were analysed and generalisations of the stability function proposed that address them were proposed. Experiments on standard datasets show, that the proposed generalisations are comparable with the other state-of-the-art methods and are beneficial in image retrieval.

Computation cost improvements

The optimisation of the running time of the baseline union-find based algorithm [14] are focused on reduction of the memory footprint and improvement of the "cache locality" of internal structures. *Suballocation* allows to reduce the costs of the system/library allocator by allocating higher number of small structures at once in one location. The *region recycling*, re-uses the region structures of terminated regions without stable thresholds and thus reduces the maximum number of concurrently used regions. The *small regions* improvement reduces the cost of small region construction by keeping limited statistics until region reaches minimal usable size. The *histogram integration* integrates the bins with sorted pixels of the image into the union find structure.

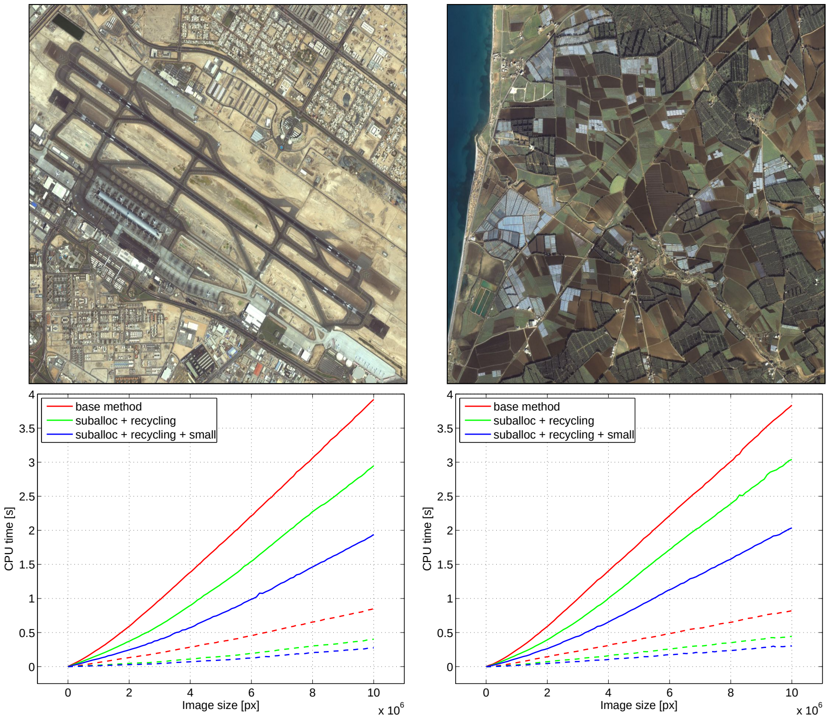
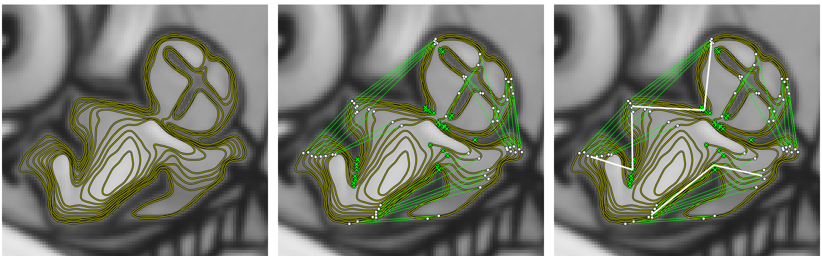


Figure 3: Example images and comparison of improvement of by suballocation, region recycling and small region structures optimisations. Solid line – overall output time, dashed line the time for output of the regions.

The set of improvements of MSER detector that reduces the number of allocations and size of memory footprint were proposed and evaluated. A significant reduction of computational time was achieved and *almost linear* asymptotic complexity was observed, comparable to the recent state-of-the-art implementation [19].

5 Stable Affine Frames

The stability of MSERs is influenced by the stability of the whole connected component. A significantly growing part of the region often lowers the stability measure. To overcome this, the ability to match parts of extremal regions between adjacent intensities in an affine covariant way has to be developed. We have introduced the notion of extended extremal region boundaries as an efficient way of enumerating all parts of discrete isophotes and an affine covariant construction of the sequences of affine frames that represent the evolution of a part of the contour over a range of intensities (see Figure 4).



(a) (b) (c)

Figure 4: Example of a stable affine frame construction: (a) each 10^{th} isophote on a part of an image, (b) entry and exit points (white) and the farthest point (green cross) from a bitangent (green lines) constructed on isophotes, (c) SAFs; white lines connecting points $(1,0)^\top$, $(0,0)^\top$ and $(0,1)^\top$ in the frame coordinate system.

A novel concept of SAFs relaxes the requirement of a stable extremal region and enables the extraction of the *stable parts* of the boundary of an extremal region. We have shown experimentally that SAFs have a repeatability comparable to the best affine covariant detectors [16] and consistently produce a significantly higher number of features per image. Overall, SAFs provide a strong alternative to MSERs (combined with local affine frame constructions) in applications where the longer running time is not an issue.

a) b) c)

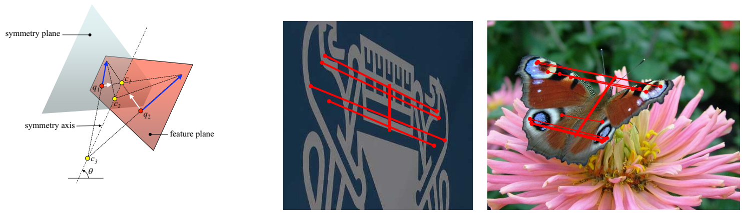
Figure 5: Mirrored LAF constructions: a), b) local affine frames and their matching mirrored versions. c) the corresponding patches on the butterfly image. Note residual differences due to different background of normalised patches and their matching mirrored counterparts.

6 Epipolar Geometry from Two LAF Pairs

The full use of the affine covariant features is as important as their repeatable detection. We have shown that by exploiting all available local geometry information in a RANSAC framework, significant speed-up of the geometry estimation and verification can be achieved. The 2LAF-LO-RANSAC algorithm requires lower number of samples, local optimisations, and significantly lower number of iterations than the seven point LO-RANSAC with degeneracy test.

7 Efficient Symmetry Detection

Symmetry is a visual and physical phenomenon, occurring both naturally and in manufactured artifacts and architecture. If two symmetric image regions are detected it is likely that these regions are related in the real world, and there is a good chance that they belong to the same object. We build on the results of Loy *et al*. [13] and Cornelius *et al*. [5] that illustrated the effectiveness of symmetry detection using local features. We take the concept proposed by [5] and improve this by removing the need to form feature quadruplets, and thus solve symmetry detection under perspective in a cleaner and more efficient manner. That is, we show how to derive a unique symmetry hypothesis from a single pair of affine covariant symmetric features, with the added challenge of an unknown perspective distortion. To allow matching of bilaterally symmetric LAFs (see Figure 5), a mirrored version of each frame and its descriptor is computed. When pairs of reflected frames have been obtained by fast matching, an axis of symmetry is calculated for all pairs where possible. A LAF is defined by three



a) b) c)

Figure 6: a) A reflective pair of LAFs (q_1, q_2) and the axis of symmetry around which the pair is symmetric. The intersections of the three lines defined by the three points in the LAFs all lie on the axis of symmetry (c_1, c_2, c_3). b),c) Symmetries discovered from the matching LAF pairs with axes of symmetry consistent with all hypothesis.

points and three lines can be obtained by connecting these points. For a symmetric pair of frames, the three pairs of corresponding lines will intersect on the axis of symmetry. This means that the axis of symmetry can be estimated from a single pair of matching LAFs (see Figure 6) if intersections c_1, c_2, and c_3 exist. Hypotheses for axes are counted and the ones with sufficient number of votes returned.

A detection of very discriminant image structures – bilateral symmetries was proposed by exploiting the local geometry of affine covariant features. In contrast to the previous work no Hough transform, is used. Hypotheses are generated from only one corresponding reflected pair of affine frames and vote for a single axis of symmetry. The complexity of the proposed algorithm is $n \log(n)$, where n is the number of affine frames detected in the image allowing in practise a near real-time performance.

8 Efficient Representation of Local Geometry

So far, we have focused on the detection and use of affine covariant local features in visual correspondence problem, ignoring memory requirements for storing the local geometry. Local geometry represented by an affine covariant ellipse is noticeably bigger than the size of tf-idf weights and labels when stored in a naive way thus becoming a significant factor determining the limits of a method. The proposed discretized local geometry representation is learnt by minimisation of the expected reprojection error after the ellipses are used to hypothesise an affine transformation. This is the optimal cost function, since the local feature geometry is stored exactly for this purpose. To reduce the memory requirements of storing the geometric information, we aim at representing a set of similar elliptical regions by a good approximation. Using simple manipulations we show that the reprojection error integrated over unit circle is a monotonic function of the Frobenius norm of the error matrix, *i.e*. the residual misalignment. Having this error measure a k-means algorithm is employed to find the ellipse cluster centres, see Figure 7, that are analogous to geometric "vocabulary".

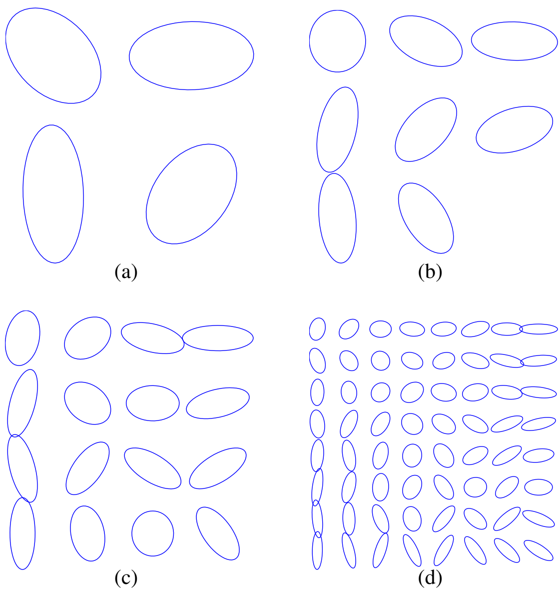


Figure 7: Learnt geometric "vocabularies". Examples of ellipse prototypes (with scale removed) for $K = 4, 8, 16$ and 64.

Experiments (*c.f*. Table 1) has shown that with 24 bits (method S0E8 - 8bits for ellipse shape, 16bits for coordinates) representing position, elliptical shape and scale (*i.e*. affine transformation modulo rotation) of each feature, the image retrieval performance is almost as good as with exact representation of local geometry. The representation naturally incorporates the gravity vector assumption that if used consistently in all stages of image retrieval from feature description to spatial verification, improves the retrieval performance.

Method	Oxford5K vocab.		Paris/Other*	
	Ox5K	Ox105K	Ox5K	Ox105K
S0E8	0.788	0.725	0.634	0.574
S0E8+SA	0.846	0.779	0.725	0.652
S0E8+QE	0.901	0.856	0.784	0.728
S0E8+SA+QE	**0.916**	**0.885**	**0.822**	**0.772**
Oxford	0.653	0.565	0.460	0.385
Oxford+QE	0.801	0.708	0.654	0.562
Oxford+SA+QE	0.825	0.719	0.718	0.605
INRIA	-	-	0.547*	-
INRIA TR	-	-	0.610*	-

Table 1: Comparison of the proposed S0E8 method with state-of-the-art methods: Oxford [23], INRIA and INRIA TR, are the best results on Oxford5K dataset in [9] resp. [10]. QE - query expansion, SA - soft assignment. , *please note that a different vocabulary was used.

9 Conclusions

In this thesis, different aspects of the visual correspondence problem were studied. We have analysed the MSER detector, proposed solution to the blur and discretisation issues, and by reducing memory footprint achieved two-fold speed up with respect to the baseline implementation. A novel detector of SAFs was proposed. The advantage of the affine over point-to-point correspondences was exploited in epipolar geometry estimation and symmetry detection. Finally an efficient representation of the local geometry was proposed and demonstrated in a large-scale retrieval system.

References

[1] V. Caselles, B. Coll, and J.M. Morel. Topographic maps and local contrast changes in natural images. *IJCV*, 33(1):5–27, 1999.

[2] O. Chum, J. Matas, and J. Kittler. Locally optimized ransac. In *Proc. of the DAGM*, pages 236–243, 2003.

[3] O. Chum, J. Matas, and Š. Obdržálek. Enhancing RANSAC by generalized model optimization. In *Proc. of the ACCV*, volume 2, pages 812–817, January 2004.

[4] O. Chum, J. Philbin, J. Sivic, M. Isard, and A. Zisserman. Total recall: Automatic query expansion with a generative feature model for object retrieval. In *ICCV*, 2007.

[5] H. Cornelius and G. Loy. Detecting bilateral symmetry in perspective. In *Proc of 5th Workshop on Perceptual Organisation in Computer Vision (POCV)*, pages 191–191. IEEE, 2006.

[6] M. Donoser and H. Bischof. 3d segmentation by maximally stable volumes (msvs). In *ICPR '06: Proceedings of the 18th International Conference on Pattern Recognition*, pages 63–66, 2006.

[7] M. Donoser and H. Bischof. Efficient maximally stable extremal region (mser) tracking. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 553–560, 2006.

[8] P. Forssén. Maximally stable colour regions for recognition and matching. In *IEEE Conference on Computer Vision and Pattern Recognition*, 2007.

[9] H. Jegou, M. Douze, and C. Schmid. Hamming embedding and weak geometric consistency for large scale image search. In *ECCV*, 2008.

[10] H. Jegou, M. Douze, and C. Schmid. Hamming embedding and weak geometry consistency for large scale image search - extended version. Technical report, INRIA, 2008.

[11] Y. Lamdan, J.T. Schwartz, and H.J. Wolfson. Object recognition by affine invariant matching. In *CVPR*, pages 335–344, 1988.

[12] D.G. Lowe. Distinctive image features from scale-invariant keypoints. *IJCV*, 60(2):91–110, 2004.

[13] G. Loy and J.O. Eklundh. Detecting symmetry and symmetric constellations of features. In *ECCV*, 2006.

[14] J. Matas, O. Chum, M. Urban, and T. Pajdla. Robust wide baseline stereo from maximally stable extremal regions. In *BMVC*, pages 384–393, 2002.

[15] J. Matas and K. Zimmermann. A new class of learnable detectors for categorisation. *Image Analysis*, pages 541–550, 2005.

[16] K. Mikolajczyk, T. Tuytelaars, C. Schmid, A. Zisserman, J. Matas, F. Schaffalitzky, T. Kadir, and L. Van Gool. A comparison of affine region detectors. *IJCV*, 65(1-2):43–72, 2005.

[17] E. Murphy-Chutorian and M. M. Trivedi. N-tree disjoint-set forests for maximally stable extremal regions. In *BMVC*, pages 739–748, 2006.

[18] D. Nistér and H. Stewénius. Scalable recognition with a vocabulary tree. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, volume 2, pages 2161–2168, 2006.

[19] D. Nistér and H. Stewénius. Linear time maximally stable extremal regions. *Computer Vision–ECCV 2008*, pages 183–196, 2008.

[20] Š. Obdržálek and J. Matas. Object recognition using local affine frames on distinguished regions. In *BMVC*, pages 113–122, 2002.

[21] Š. Obdržálek and J. Matas. Sub-linear indexing for large scale object recognition. In *BMVC*, 2005.

[22] J. Philbin, O. Chum, M. Isard, J. Sivic, and A. Zisserman. Object retrieval with large vocabularies and fast spatial matching. In *CVPR*, 2007.

[23] J. Philbin, O. Chum, M. Isard, J. Sivic, and A. Zisserman. Lost in quantization: Improving particular object retrieval in large scale image databases. In *CVPR*, 2008.

[24] J. Sivic and A. Zisserman. Video google: A text retrieval approach to object matching in videos. In *International Conference on Computer Vision*, volume 2, pages 1470–1477, 2003.

[25] T. Tuytelaars, A. Turina, and L. Van Gool. Noncombinatorial detection of regular repetitions under perspective skew. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 25(4):418–432, 2003.

[26] T. Tuytelaars and L. Van Gool. Wide baseline stereo matching based on local, affinely invariant regions. In *BMVC*, pages 412–425, 2000.

Resumé in Czech

Problém nalezení korespondencí ve dvou různých pohledech na objekt nebo scénu je v počítačovém vidění zkoumán již dlouho. Nedávno publikované metody rozpoznávaní objektů a vyhledávaní obrázků jsou založeny na vizuálních korespondencích nalezených pomocí párování invariantních popisů lokálních příznaků. Problém nalezení vizuálních korespondencí mezi hledaným obrázkem a obrázky v databázi je formulován jako hledání geometricky konzistentních množin podobných popisů lokálních příznaků. Algoritmus pro řešení takto formulovaného problému se skládá z následujících komponent: detektor příznaků, generátor popisů, indexování a ověřování korespodencí. Metody pro realizaci každé z těchto komponent jsou známé, ale tyto mají různá omezení a nedostatky. Tato práce se zabývá analýzou a řešením těchto problémů.

Nejdříve analyzujeme některé nedostatky detektoru Maximálně stabilních extremálních oblastí. Navrhujeme rozšíření funkce pro výpočet stability, analyzujeme kritické body implementace detektoru založené na Union-Find struktuře a navrhujeme zlepšení která snižují velikost paměti potřebné k běhu a počet alokací paměti. Navrhujeme nový koncept lokální příznaků nazvaných Stabilní afinní rámce, které odstraňují potřebu nalezení stabilní oblasti a umožňují detekci stabilních částí hranice extremální oblasti.

Dále využívame vizuální korespondence afinně kovariatních příznaků k nalezení a ověření epipolární geometrie. Navržený algoritmus používá RANSAC s lokální optimalizací a detekcí degenerovaných konfigurací k nalezení epipolární geometrie ze dvou afinních korespondencí. Algoritmus umožňuje rychlejší ověření globálních geometrických omezení, často poslední z kroků při rozpoznávání objektů, hledání korespondencí ze dvou různých pohledů a vyhledávání obrázků.

Dále se zabýváme hledáním symetrie, jevem jenž se objevuje v přírodě i na člověkem vytvořených objektech a architektuře. Výskyt symetrie v obrázku objektu je často nenáhodný a velice charakteristický. Existující algoritmus pro nalezení bilaterální symetrie pomocí lokálních příznaků je rozšířen tak, že se k hledání hypotézy osy symetrie výužívá pouze jediná korespondence afinně kovariatních lokálních příznaků.

Nedávný rozvoj systémů pro vyhledávání ve velkých databázích obrázků s sebou přinesl nový problém omezené kapacity paměti pri ukládání geometrie lokálních příznaků. V práci navrhujeme novou reprezentaci lokální geometrie založenou na kompresi a ukazujeme zásadní úsporu pamětových nároku v aplikaci pro vyhledávání specifických objektů.

Studované problémy jsou doprovázené důkladným experimentálním ověřením na standardních souborech obrázků pro porovnávání detektorů a v aplikaci pro vyhledávaní specifických objektů.

Author's publications

Journal Papers:

[P00] J. Čech, J. Matas, and M. Perďoch. Efficient Sequential Correspondence Selection by Cosegmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 32(9):1568–1581, September 2010. Authorship: 50-30-20

Conference Papers:

[P01] M. Perďoch, J. Matas, and O. Chum. Epipolar Geometry from Two Correspondences. In *ICPR 2006: Proceedings of the 18th International Conference on Pattern Recognition*, pages 215–220, Los Alamitos, 2006. IEEE Computer Society Press. Authorship: 34-33-33.

[P02] M. Perďoch, J. Matas, and Š. Obdržálek. Stable Affine Frames on Isophotes. In *ICCV 2007: Proceedings of Eleventh IEEE International Conference on Computer Vision*, pages –, Madison, 2007. Omnipress. Authorship: 50-30-20.

[P03] H. Cornelius, M. Perďoch, J. Matas, and G. Loy. Efficient Symmetry Detection Using Local Affine Frames. In *SCIA 2007: Proceedings of 15th Scandinavian Conference on Image Analysis*, pages 152–161, Heidelberg, 2007. Springer. Authorship: 30-30-20-20.

[P04] J. Čech, J. Matas, and M. Perďoch. Efficient Sequential Correspondence Selection by Cosegmentation. In *CVPR 2008: Proceedings of the 2008 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pages 1020–1027, Medison, 2008. Omnipress. Authorship: 50-30-20.

[P05] M. Perďoch, O. Chum, and J. Matas. Efficient Representation of Local Geometry for Large Scale Object Retrieval. In *CVPR 2009: Proceedings of the 2009 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pages 9–16, Madison, 2009. Omnipress. Authorship: 34-33-33.

[P06] A. Mikulík, J. Matas, M. Perďoch, and O. Chum. Construction of Precise Local Affine Frames. In *ICPR'2010: Proceedings of the 20th International Conference on Pattern Recognition*, pages 3565–3569, Los Alamitos, 2010. IEEE Computer Society Press. Authorship: 25-25-25-25.

[P07] P. Doubek, J. Matas, M. Perďoch, and O. Chum. Image Matching and Retrieval by Repetitive Patterns. In *ICPR'2010: Proceedings of the 20th International Conference on Pattern Recognition*, pages 3195–3198, Los Alamitos, 2010. IEEE Computer Society Press. Authorship: 25-25-25-25.

Journal Papers not Directly Related to the Thesis:

[P08] C. Lemaitre, M. Perďoch, A. Rahmoune, J. Matas, and J. Miteran. Detection and matching of curvilinear structures. *Pattern recognition*, 44(7):1514–1527, July 2011. Authorship: 20-20-20-20-20.

Conference Papers not Directly Related to the Thesis:

[P09] P. Doubek, M. Perďoch, J. Matas, and J. Šochman. Mobile Mapping of Vertical Traffic Infrastructure. In *CVWW 2008: Proceedings of the 13th Computer Vision Winter Workshop*, pages 115–122, Ljubljana, 2008. Slovenian Pattern Recognition Society. Authorship: 70-15-10-5.

[P10] Š. Obdržálek, M. Perďoch, and J. Matas. Dense Linear-Time Correspondences for Tracking. In *Proceedings of Workshop on Visual Localization for Mobile Platforms held in conjunction with CVPR 2008*, pages 1302–1309, Piscataway, 2008. IEEE. Authorship: 40-30-30.

[P11] O. Chum, M. Perďoch, and J. Matas. Geometric min-Hashing: Finding a (thick) needle in a haystack. In *CVPR 2009: Proceedings of the 2009 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pages 17–24, Madison, 2009. Omnipress. Authorship: 34-33-33.

[P12] A. Mikulík, M. Perďoch, O. Chum, and J. Matas. Learning a Fine Vocabulary. In *Computer Vision - ECCV 2010, 11th European Conference on Computer Vision, Proceedings, Part III*, pages 1–14, Heidelberg, 2010. Springer. Authorship: 25-25-25-25.

Citations of author's work

Journals:

[C01] Timo Dickscheid, Falko Schindler, and Wolfgang Foerstner. Coding Images with Local Features. *International Journal of Computer Vision*, 94(2):154–174, SEP 2011.

[C02] Bernd Girod, Vijay Chandrasekhar, David M. Chen, Ngai-Man Cheung, Radek Grzeszczuk, Yuriy Reznik, Gabriel Takacs, Sam S. Tsai, and Ramakrishna Vedantham. Mobile Visual Search [Linking the virtual and physical worlds]. *IEEE Signal Processing Magazine*, 28(4):61–76, JUL 2011.

[C03] Eagle S. Jones and Stefano Soatto. Visual-inertial navigation, mapping and localization: A scalable real-time causal approach. *International Journal of Robotics Research*, 30(4):407–430, APR 2011.

[C04] Shiliang Zhang, Qi Tian, Gang Hua, Wengang Zhou, Qingming Huang, Houqiang Li, and Wen Gao. Modeling spatial and semantic cues for large-scale near-duplicated image retrieval. *Computer Vision and Image Understanding*, 115(3, SI):403–414, MAR 2011.

[C05] Yannis Kalantidis, Giorgos Tolias, Yannis Avrithis, Marios Phinikettos, Evaggelos Spyrou, Phivos Mylonas, and Stefanos Kollias. VIRaL: Visual Image Retrieval and Localization. *Multimedia Tools And Applications*, 51(2, SI):555–592, JAN 2011.

[C06] Herve Jegou, Matthijs Douze, and Cordelia Schmid. Product Quantization for Nearest Neighbor Search. *IEEE Transactions on Pattern Analysis And Machine Intelligence*, 33(1):117–128, JAN 2011.

[C07] Jan-Michael Frahm, Marc Pollefeys, Svetlana Lazebnik, David Gallup, Brian Clipp, Rahul Raguram, Changchang Wu, Christopher Zach, and Tim Johnson. Fast robust large-scale mapping from video and internet photo collections. *ISPRS Journal of Photogrammetry and Remote Sensing*, 65(6):538–549, NOV 2010.

[C08] Amit Bermanis, Amir Averbuch, and Yosi Keller. 3-D Symmetry Detection and Analysis Using the Pseudo-polar Fourier Transform. *International Journal of Computer Vision*, 90(2):166–182, NOV 2010.

[C09] Seungkyu Lee and Yanxi Liu. Skewed Rotation Symmetry Group Detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 32(9):1659–1672, SEP 2010.

[C10] Noah Snavely, Ian Simon, Michael Goesele, Richard Szeliski, and Steven M. Seitz. Scene Reconstruction and Visualization from Community Photo Collections. *Proceedings of the IEEE*, 98(8, Si):1370–1390, Aug 2010.

[C11] Hao Ma, Jianke Zhu, Michael Rung-Tsong Lyu, and Irwin King. Bridging The Semantic Gap Between Image Contents and Tags. *IEEE Transactions on Multimedia*, 12(5):462–473, AUG 2010.

[C12] Jasna Maver. Self-Similarity and Points of Interest. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 32(7):1211–1226, JUL 2010.

[C13] Michal Sofka and Charles V. Stewart. Location registration and recognition (LRR) for serial analysis of nodules in lung CT scans. *Medical Image Analysis*, 14(3):407–428, JUN 2010.

[C14] Liran Goshen and Ilan Shimshoni. Balanced Exploration and Exploitation Model search for efficient epipolar geometry estimation. *IEEE Transactions On Pattern Analysis And Machine Intelligence*, 30(7):1230–1242, JUL 2008.

Conference Proceedings:

[C15] Sam S. Tsai, David Chen, Gabriel Takacs, Vijay Chandrasekhar, Ramakrishna Vedantham, Radek Grzeszczuk, and Bernd Girod. Fast Geometric Re-Ranking for Image-Based Retrieval. In *IEEE International Conference on Image Processing ICIP*, pages 1029–1032, 2010. Hong Kong, China, SEP 26-29, 2010.

[C16] Chenbo Shi, Guijin Wang, Xinggang Lin, Yongming Wang, Chao Liao, and Quan Miao. Topology Based Affine Invariant Descriptor For Msers. In *IEEE International Conference on Image Processing ICIP*, pages 133–136, 2010. Hong Kong, China, SEP 26-29, 2010.

[C17] Afifa Dahmane, Slimane Larabi, and Chabane Djeraba. Detection And Analysis Of Symmetrical Parts On Face For Head Pose Estimation. In *IEEE International Conference on Image Processing ICIP*, pages 3249–3252, 2010. Hong Kong, China, SEP 26-29, 2010.

[C18] Yong Jae Lee and Kristen Grauman. Collect-Cut: Segmentation with Top-Down Cues Discovered in Multi-Object Images. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 3185–3192, 2010. San Francisco, CA, JUN 13-18, 2010.

[C19] Shai Bagon, Ori Brostovski, Meirav Galun, and Michal Irani. Detecting and Sketching the Common. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 33–40, 2010. San Francisco, CA, JUN 13-18, 2010.

[C20] Herve Jegou, Matthijs Douze, Cordelia Schmid, and Patrick Perez. Aggregating local descriptors into a compact image representation. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 3304–3311, 2010. San Francisco, CA, JUN 13-18, 2010.

[C21] Yang Cao, Changhu Wang, Zhiwei Li, Liqing Zhang, and Lei Zhang. Spatial-Bag-of-Features. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 3352–3359, 2010. San Francisco, CA, JUN 13-18, 2010.

[C22] Kyle Heath, Natasha Gelfand, Maks Ovsjanikov, Mridul Aanjaneya, and Leonidas J. Guibas. Image Webs: Computing and Exploiting Connectivity in Image Collections. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 3432–3439, 2010. San Francisco, CA, JUN 13-18, 2010.

[C23] Wenying Wang, Dongming Zhang, Yongdong Zhang, Jintao Li, and Xiaoguang Gu. Parallel Spatial Matching For Object Retrieval Implemented On GPU. In *2010 IEEE International Conference On Multimedia And Expo (ICME 2010)*, pages 890–895, 2010. Singapore, JUL 19-23, 2010.

[C24] Andrzej Sluzek and Mariusz Paradowski. Real-Time Retrieval of Near-Duplicate Fragments in Images and Video-Clips. In *Advanced Concepts For Intelligent Vision Systems, PT I*, volume 6474 of *Lecture Notes in Computer Science*, pages 18–29, 2010. Sydney, Australia, DEC 13-16, 2010.

[C25] Georges Baatz, Kevin Koeser, David Chen, Radek Grzeszczuk, and Marc Pollefeys. Handling Urban Location Recognition as a 2D Homothetic Problem. In *Computer Vision - ECCV 2010, PT VI*, volume 6316 of *Lecture Notes in Computer Science*, pages 266–279, 2010. Heraklion, Greece, SEP 05-11, 2010.

[C26] Zhe Lin and Jonathan Brandt. A Local Bag-of-Features Model for Large-Scale Object Retrieval. In *Computer Vision - ECCV 2010, PT VI*, volume 6316 of *Lecture Notes in Computer Science*, pages 294–308, 2010. Heraklion, Greece, SEP 05-11, 2010.

[C27] James Philbin, Michael Isard, Josef Sivic, and Andrew Zisserman. Descriptor Learning for Efficient Retrieval. In *Computer Vision-ECCV 2010, PT III*, volume

6313 of *Lecture Notes in Computer Science*, pages 677–691, 2010. Heraklion, Greece, SEP 05-11, 2010.

[C28] C. Lawrence Zitnick. Binary Coherent Edge Descriptors. In *Computer Vision-ECCV 2010, PT II*, volume 6312 of *Lecture Notes in Computer Science*, pages 170–182, 2010. Heraklion, Greece, SEP 05-11, 2010.

[C29] Yunpeng Li, Noah Snavely, and Daniel P. Huttenlocher. Location Recognition Using Prioritized Feature Matching. In *Computer Vision-ECCV 2010, PT II*, volume 6312 of *Lecture Notes in Computer Science*, pages 791–804, 2010. Heraklion, Greece, SEP 05-11, 2010.

[C30] David C. Lee, Qifa Ke, and Michael Isard. Partition Min-Hash for Partial Duplicate Image Discovery. In *Computer Vision-ECCV 2010, PT I*, volume 6311 of *Lecture Notes in Computer Science*, pages 648–662, 2010. Heraklion, Greece, SEP 05-11, 2010.

[C31] Nobuyuki Morioka and Shin'ichi Satoh. Building Compact Local Pairwise Codebook with Joint Feature Space Clustering. In *Computer Vision-ECCV 2010, PT I*, volume 6311 of *Lecture Notes in Computer Science*, pages 692–705, 2010. Heraklion, Greece, SEP 05-11, 2010.

[C32] Sandra Ebert, Diane Larlus, and Bernt Schiele. Extracting Structures in Image Collections for Object Recognition. In *Computer Vision-ECCV 2010, PT I*, volume 6311 of *Lecture Notes in Computer Science*, pages 720–733, 2010. Heraklion, Greece, SEP 05-11, 2010.

[C33] Jan Knopp, Josef Sivic, and Tomas Pajdla. Avoiding Confusing Features in Place Recognition. In *Computer Vision-ECCV 2010, PT I*, volume 6311 of *Lecture Notes in Computer Science*, pages 748–761, 2010. Heraklion, Greece, SEP 05-11, 2010.

[C34] Wenying Wang, Dongming Zhang, Yongdong Zhang, and Jintao Li. Fast And Robust Spatial Matching For Object Retrieval. In *2010 IEEE International Conference on Acoustics, Speech, and Signal Processing*, pages 1238–1241, 2010. Dallas, TX, MAR 14-19, 2010.

[C35] Juho Kannala, Esa Rahtu, Sami S. Brandt, and Janne Heikkila. Dense and Deformable Motion Segmentation for Wide Baseline Images. In *Proceedings of the 16th Scandinavian Conference on Image Analysis*, volume 5575 of *Lecture Notes in Computer Science*, pages 379–389, 2009. Oslo, Norway, JUN 15-18, 2009.

[C36] Friederike Schwarzbach. Suitability of Different LIDAR Data Sets for 3D Mapping of the Road Environment. *Photogrammetrie Fernerkundung Geoinformation*, (2):117–127, 2009.

[C37] Kevin Koeser, Christian Beder, and Reinhard Koch. Conjugate rotation: Parameterization and estimation from an affine feature correspondence. In *2008 IEEE Conference On Computer Vision And Pattern Recognition*, pages 3528–3535, 2008. Anchorage, AK, JUN 23-28, 2008.