

# Construction of Precise Local Affine Frames

Andrej Mikulik, Jiri Matas, Michal Perdoch, Ondrej Chum  
Center for Machine Perception  
Czech Technical University in Prague  
Czech Republic  
e-mail: mikulik@cmp.felk.cvut.cz

**Abstract**—We propose a novel method for the refinement of Maximally Stable Extremal Region (MSER) boundaries to sub-pixel precision by taking into account the intensity function in the  $2 \times 2$  neighborhood of the contour points. The proposed method improves the repeatability and precision of Local Affine Frames (LAFs) constructed on extremal regions. Additionally, we propose a novel method for detection of local curvature extrema on the refined contour.

Experimental evaluation on publicly available datasets shows that matching with the modified LAFs leads to a higher number of correspondences and a higher inlier ratio in more than 80% of the test image pairs. Since the processing time of the contour refinement is negligible, there is no reason not to include the algorithms as a standard part of the MSER detector and LAF constructions.

**Keywords**-discretized contour, contour refinement, curvature estimation, curvature extrema

## I. INTRODUCTION

The Maximally Stable Extremal Region (MSER) detector [1] has been shown to be a reliable state-of-the-art region detector [2]. The detected regions have high repeatability, they are robust to photometric transformations and viewpoint changes and in many types of scenes outperforms other affine covariant region detectors [2].

Since the boundary of an MSER region is defined by a set of pixels, the contour is affected by rasterization. We propose a novel method to the refinement of the discretized contour based on the gradient of image intensity. The key idea is to exploit the fact that the boundary of MSER is a discretized isophote of a smooth intensity function, thus the local gradient is assumed to be constant. The method estimates subpixel position of the isophote examining pixel values in the local neighborhood of the contour. The refined contour facilitates improvement in detection of contours primitives like the inflection point or curvature extrema. In contrast, reference approaches [3] and [4] use curvature scale-space to suppress discretization noise. The main drawback of the reference method is the loss of the precision (see Figure 1 – green line).

Local Affine Frames (LAFs) [5] define measurement regions [1] on affine covariant primitives extracted from detected regions. The frames, consisting of ordered triplets of points, are more useful than affine covariant regions as they directly facilitate affine invariant description of



Figure 1. A comparison of three region boundaries. White — the original MSER boundary, Green — smoothed with the Gaussian filter [3], Red — contour refinement (the proposed approach)

the image signal without any further processing such as detection of dominant gradient directions.

Discretization effects on the original MSER boundary negatively influence the precision of the created LAFs. The goal of this work is to show that a simple refinement of the contour and thus more precise localization of affine covariant primitives leads to a higher repeatability of LAF constructions and improves recognition performance.

An experimental evaluation was conducted on two publicly available datasets (ZuBuD [6] and Mikolajczyk's [2]) to show the performance of the proposed method in real applications. In the evaluation, we focused on two aspects, wide-baseline matching performance and geometric precision.

The rest of the paper is structured as follows. An overview of the reference method with the proposed approaches are described in Section II. Experimental validation follows in Section III and the paper is concluded in Section IV.

## II. PRECISE LAF CONSTRUCTION ON THE REFINED MSER CONTOUR

A local affine frame (LAF) [3] is a coordinate system constructed by combining affine-covariant primitives detected on an MSER which constrains all six degrees of freedom of an affine transformation. A local patch around a LAF is normalized to a canonical coordinate system and finally

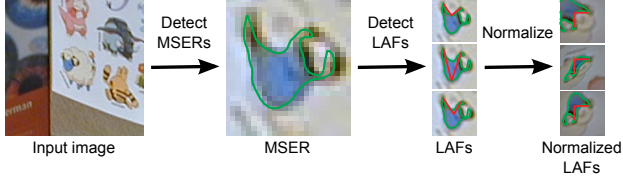


Figure 2. LAF construction on MSER contour.

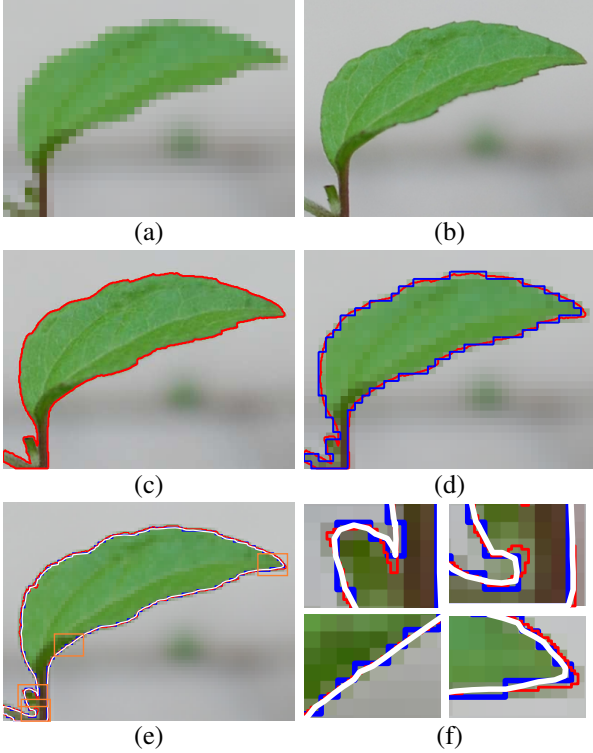


Figure 3. Discretization effects on an MSER region. (a) Input image for the MSER detector. (b) The same image, with 10 times higher resolution used as ground truth. (c) The detected MSER on the high-resolution image. (d) The detected MSER (blue line) on input image and the “ground truth” projected from the high-resolution image. (e) White line - contour refined by the proposed method. (f) Close-ups.

a descriptor (e.g. SIFT) is computed (see Figure 2). The affine covariant primitives constructed on MSER contour are, e.g., local curvature extrema, inflection points, or center of gravity. Localization of the primitives, especially on smaller regions, is severely affected by image rasterization (see Figure 3).

MSERs are returned by the detector as 4-connected sets of image pixels. The outer contour of the region  $C = (\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_n)$ ,  $\mathbf{a}_i \in \mathbb{R}^2$  is a Jordan curve through the corners of image pixels  $\mathbf{a}_i = (x, y)$ ,  $\forall i : \|\mathbf{a}_i - \mathbf{a}_{(i \bmod n)+1}\| = 1$ . Contours around holes in the MSERs are treated separately and in the same way as the outer contours. In the reference methods [4] and [3], the contour is smoothed with a Gaussian filter to avoid discretization effects:

$$\mathbf{a}'_i = \frac{\sum_t \mathbf{a}_t G_{0,\sigma}(d(\mathbf{a}_i - \mathbf{a}_t))}{\sum_t G_{0,\sigma}(d(\mathbf{a}_i - \mathbf{a}_t))}, \quad (1)$$

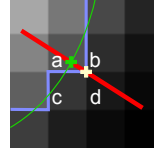


Figure 4. Shifting the vertex on the original contour (yellow cross) to a new position (green) based on the intensity of pixels  $a, b, c, d$ . Red line - gradient direction, green line - the isophote, blue line - the original contour.

where  $\mathbf{a}'_i$  is position of contour vertex  $i$ ,  $d(\mathbf{a}_i, \mathbf{a}_t)$  is the distance along contour between vertexes  $i$  and  $t$ , and  $G_{0,\sigma}(\bullet)$  is a Gaussian PDF with zero mean and standard deviation  $\sigma$ . To preserve scale invariance of the construction,  $\sigma$  must be proportional to the square root of the region area:

$$\sigma = \max(\sqrt{|\Omega|}/k, 1), \quad (2)$$

where  $|\Omega|$  is size of the region (number of region pixels), and  $k$  is parameter controlling the smoothing. This step is not affine covariant but is fast and sufficiently suppresses the rasterization effect and it had a positive impact on the recognition rate in the object recognition experiment [3].

The main problem of the reference method is the choice of the level of smoothing (setting of the parameter  $k$ ) a trade-off between preserving local structures and suppression of rasterization effects. To avoid this trade-off we propose a novel method for contour refinement. Instead of warping the outer contour around pixels with the intensity at or above the intensity threshold (as done in standard MSER), the proposed method shifts the location of contour point to the estimated location of the isophote exactly at the intensity threshold. The desired output of the new contour refinement algorithm is contour  $C'$  with following properties:

- $C' = (\mathbf{a}'_1, \mathbf{a}'_2, \dots, \mathbf{a}'_n)$ ,  $\mathbf{a}_i \in \mathbb{R}^2$ ,
- $\forall i : \|\mathbf{a}_i - \mathbf{a}'_i\|_\infty < 0.5$ , where  $\|\bullet\|_\infty$  denotes  $L_\infty$  norm,
- $C'$  is Jordan curve,
- $|C'| = |C|$ .

In order to estimate the isophote, the intensity function in four neighboring pixels of each contour vertex is examined (see Figure 4). The only assumption is a constant gradient in a small neighborhood. First, a gradient of intensity function is estimated using Roberts operator for each vertex. This is

---

#### Algorithm 1 Subpixel refinement

---

Input: region contour  $C = (\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_n)$ , image  $I$ , intensity threshold  $t$   
Output: refined contour  $C' = (\mathbf{a}'_1, \mathbf{a}'_2, \dots, \mathbf{a}'_n)$

For each contour vertex  $\mathbf{a} \in C$  estimate refined position  $\mathbf{a}' \in C'$  in the following way:

- 1) Approximate the gradient  $\nabla I(\mathbf{a}) \left( \frac{\partial I(\mathbf{a})}{\partial x}, \frac{\partial I(\mathbf{a})}{\partial y} \right)$  of the image intensity function from four neighboring pixels  $p_1, \dots, p_4$  of vertex  $\mathbf{a}$  using Roberts operator [7].
  - 2) Project center points of pixels  $p_1, \dots, p_4$  to  $p'_1, \dots, p'_4$  with orthogonal projection to line passing through vertex  $\mathbf{a}$  in the gradient direction.
  - 3) Estimate the new position  $\mathbf{a}'$  of the vertex  $\mathbf{a}$  for threshold  $t$  by linear regression. The regressors are pixel intensities  $I(p_i)$ , and  $y_i = \|p'_i - \mathbf{a}\|$  are the regressands.
-

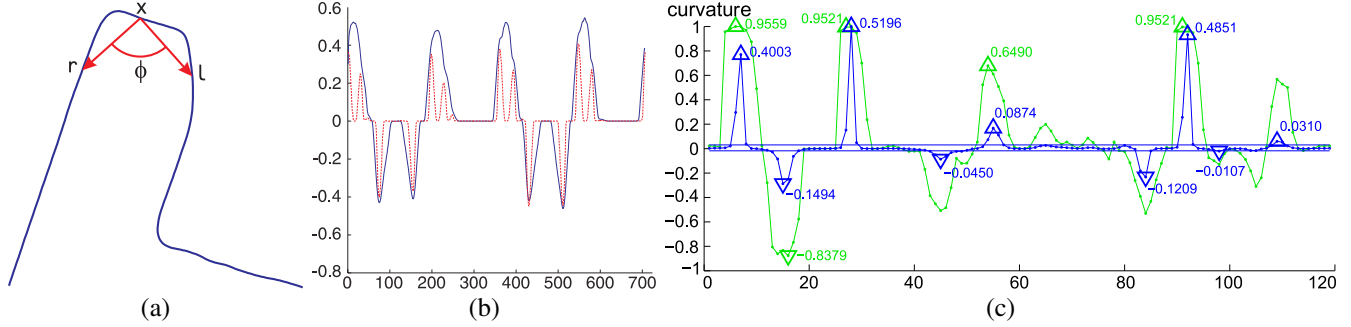


Figure 5. (a) Curvature estimation. (b) Curvature computed for two different values of  $a$ ,  $a = 0.5$  (blue line) and  $a = 0.2$  (red line). (c) Blue line – the curvature at each point of the refined boundary. Detected extrema are marked by triangles. Green line – the reference method. Note the better localization (narrower peaks) of the extrema on the blue line. The straight blue lines near zero curvature depict detection thresholds for curvature extrema.

accomplished with two  $2 \times 2$  kernels:

$$\begin{pmatrix} 1 & -1 \\ 1 & -1 \end{pmatrix}, \text{ and } \begin{pmatrix} 1 & 1 \\ -1 & -1 \end{pmatrix}.$$

Second, the new position of the isophote is estimated by linear regression. The contour vertex is then shifted to the approximated nearest position on an isophote.

In some pathological cases, like very narrow part of region or a very sharp corner, it is possible that the new estimated position of the vertex will be further than half pixel away from the original position. This is consequence of the fact that the assumptions about constant gradient in the neighborhood is not satisfied. The whole contour refinement procedure is summarized in Algorithm 1.

Contour refinement is the first step of the proposed local curvature extrema extraction. In the reference work [3] the curvature is defined as follows: for each vertex  $\mathbf{x}$ , two segments  $l$  and  $r$  of length  $a$  are cast on opposite directions along the polygon boundary (see Figure 5).

The curvature  $\kappa$  is estimated from the cosine of the angle  $\phi$ :

$$\cos \phi = \frac{l_x r_x + l_y r_y}{|l||r|},$$

as follows:

$$\kappa = s \frac{1 + \cos \phi}{2}, \quad s = \begin{cases} 1 & \text{if } l_x r_y - l_y r_x > 0 \\ -1 & \text{otherwise} \end{cases} \quad (3)$$

The curvature ranges from  $-1$  to  $1$ , equals to  $0$  for straight segments, and is negative for concave and positive for convex curvatures (see Figure 5). The curvature of neighboring vertexes is affected by the choice of the length  $a$  (Figure 5b). We observed that long spans cause undesirable artifacts on smaller regions with lower resolution. To achieve higher stability and better localization we propose a novel approach for extracting local curvature extrema. The proposed approach differs in three main points:

- a contour vertex is a local curvature extremum, if it is a pre-image of detected local extremum on a smoothed contour,

---

### Algorithm 2 Local curvature extrema detection

---

Input: region contour  $C$ ,  $P = \text{trace}(C)$

Output: curvature extrema  $E \subset P$

- 1)  $C' = \text{smooth}(C, \sigma)$

Smooth the contour with  $G_{0,\sigma}$ ,  $\sigma$  cf. Equation 2,  $k$  is set to 60 (contrary to a much milder smoothing  $k = 30$ , which is typical if it is necessary to preserve position of the contour).

- 2)  $N' = AC'$ , where  $A = (\text{chol}(\Sigma))^{-1}$

Normalize the contour by an inverse of Cholesky decomposition of the covariance matrix  $\Sigma$ .

- 3) Estimate curvature  $\kappa_i$  at each contour point  $p'_i \in \text{trace}(N')$ , using Equations 3.

- 4)  $\mu = E(K)$ ,  $\alpha = \sqrt{\frac{1}{n} \sum_{i=1}^n (\mu - \kappa_i)^2}$

Compute mean value  $\mu$  and standard deviation  $\alpha$  of estimated curvatures  $\kappa_i$ .

- 5)  $K = \{p'_i \in \text{trace}(N') \mid |\kappa_i - \mu| > 3\sigma\}$

$K$  is the set of local curvature extrema  $p'$  with curvature exceeding  $3\sigma$ .

- 6) The local extrema  $E' \subseteq K$  are selected from the candidates  $K$  by non-maxima suppression.

- 7) Points  $e \in E \subset P$  – the pre-images of  $e' \in E'$  – are the locations of the local curvature extrema on the original curve.
- 

- smoothing in the first phase is much stronger than in the reference method, since it does not affect precise localization of contour extrema,
- thanks to stronger smoothing, much shorter segments  $l$  and  $r$  are successfully used in curvature estimation (Equation 3) to achieve higher stability and suppress artifacts mentioned above.

The proposed method is summarized in Algorithm 2. For more detailed information see [8].

### III. EXPERIMENTS

The performance of the proposed approach was evaluated on the state-of-the-art wide-baseline matching approach [3] (Algorithm 3). Two publicly available datasets were used in the experiments: ZuBuD [6] and Mikolajczyk's [2] (Figure 7). Three approaches to MSER boundary processing were compared: *plain* – rough MSER contour without

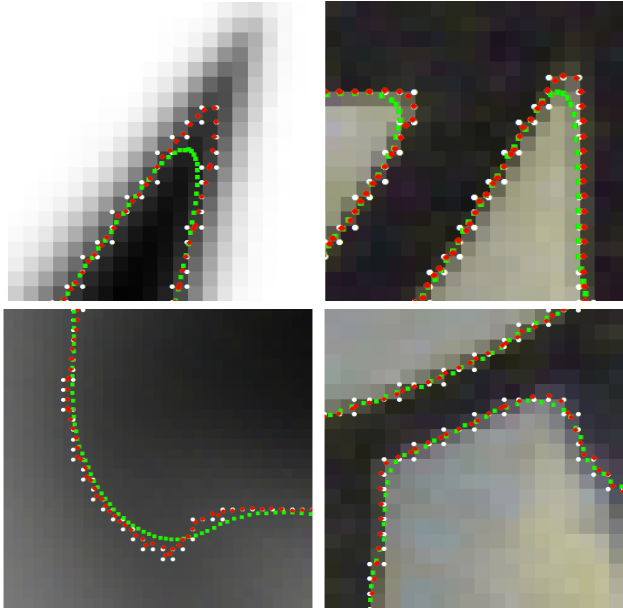


Figure 6. The comparison of three region contours. White — the original MSER boundary, Green — smoothed with the Gaussian filter [3], Red — contour refinement (the proposed approach)

further processing, *smooth* – Gaussian filtering from the reference approach [3], *refine* – the proposed method. We have focused on two aspects, repeatability and geometric precision of the proposed method.

#### A. Repeatability comparison

It has been shown [3] that the elimination of discretization effects improves repeatability. However the reference approach has some disadvantages. One of them is misplacing of contour primitives. The largest displacement is introduced to local curvature extrema (Figure 6), concavities and bitangents. Other primitives such as inflection points, or linear boundary segments gain mainly by improved curvature definition. At last, there are some primitives that are virtually unaffected by boundary processing, *e.g.* center of gravity, matrix of second moments, or orientation of gradients.

In this section, we analyze the repeatability of LAF constructions before and after contour refinement, and compare the proposed approach with the reference one.

---

#### Algorithm 3 Wide-baseline matching algorithm

---

- 1) MSERs are detected in both images.
  - 2) LAFs are constructed on MSERs with (without) contour refinement, photometrically normalized and described by descriptors.
  - 3) Tentative correspondences are established by finding mutually nearest descriptions.
  - 4) RANSAC algorithm is used to find inliers to a global model of geometric transformation – epipolar geometry or homography. (Inliers are tentative correspondences that are consistent with the global model of geometric transformation).
- 



Figure 7. Sample images from the ZuBuD [6] (top) and Mikolajczyk's datasets (bottom).

In the repeatability experiment, the number of inliers and the inlier ratio – the number of inliers divided by the number of tentative correspondences – are measured. The higher the inlier ratio, the faster the geometric verification by RANSAC [9]. The number of inliers influences the precision of the model, the probability that the model found by RANSAC is correct, and improves robustness to occlusion.

The analysis is done on a wide-baseline matching problem on the ZuBuD dataset. For all pairs formed out of the five images of each building (2010 pairs in total), we run the process of finding correspondences (Algorithm 3). Since the ground truth transformations for pairs of images are not provided with the ZuBuD dataset, the epipolar geometry with the highest number of inliers (out of all three approaches) was used to evaluate the quality of the tentative correspondences in RANSAC. Similar analysis have been done for Mikolajczyk's dataset. In this case, the ground truth homography transformations are provided with the dataset. Each of the 5 scenes contains 6 images and homographies from the first image to the others. All image pairs (25 pairs) are matched in the experiment.

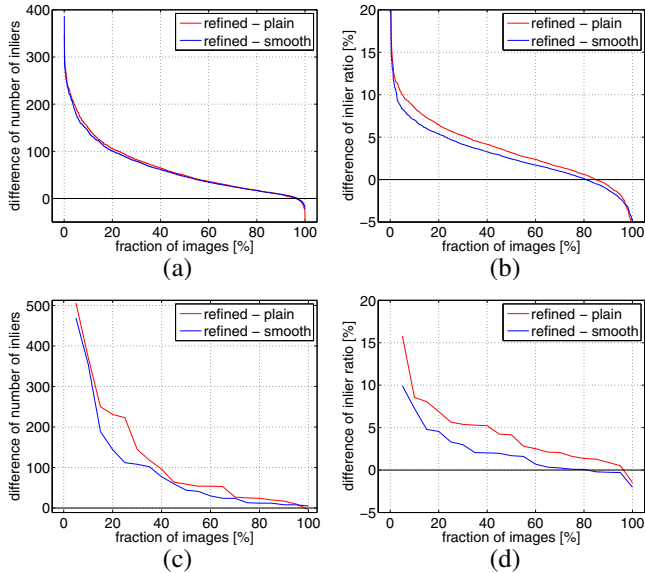
The overall gain of the proposed approach is shown in Figure 8 for the ZuBuD dataset and the Mikolajczyk's dataset. We compare the improvement ( $\Delta = new - old$ ) of the proposed method *refine* over the reference *smooth* and *plain* method (without smoothing). The proposed method gives better inlier ratio (difference is above zero) in 80% of image pairs and more inliers in 95% of image pairs.

#### B. Precision comparison

It is difficult to measure the geometric precision of LAFs directly. We decided to use geometric hashing with Local Affine Frames [10]. Each LAF is described by relative poses of other LAFs in its affine covariant neighborhood in form of a geometric hash. This representation is insensitive to a wide range of photometric changes, robust to local occlusions and computationally efficient, but sensitive to the precision of LAFs. Therefore, the performance of geometry based matching is a sensitive indicator of LAF's precision.

Tentative correspondences of LAF obtained by matching of geometric hashes are ordered by the number of votes (probability of being a correct match). It is possible to exploit this ordering in the geometric verification step using PROSAC [11] instead of RANSAC. Hence, to highlight the improvement in the matching results, inlier ratio in 500 best





Mean differences	ZuBuD		Mikolajczyk's	
	$\Delta$ inliers	$\Delta$ inlier ratio	$\Delta$ inliers	$\Delta$ inlier ratio
refine - smooth	62.5	2.79%	91.9	2.04%
refine - plain	65.1	3.56%	117.2	4.15%

Figure 8. Repeatability comparison on the ZuBuD (top) and Mikolajczyk's (bottom) datasets. (a), (c) differences in the number of inliers. (b), (d) differences in the inlier ratios.

tentative correspondences is considered.

Table I shows the results of matching of two pairs from Mikolajczyk's dataset. The proposed method results in higher number of inliers while preserving inlier ratio. For the first 500 best tentative correspondences the inlier ratio achieved by proposed method is significantly higher.

#### IV. CONCLUSIONS

A new method for MSER contour refinement that takes into account image intensity function was proposed and a localization of local curvature extrema improved. The method achieved (on two publicly available datasets) higher number of correspondences and higher inlier ratio in more than 80% of the image pairs and provides more inliers in 95% of all image pairs. Experiment with geometric hashing [10] confirms higher geometric precision of detected LAFs.

#### ACKNOWLEDGEMENTS

The authors were supported by EC projects FP7-ICT-247022 MASH, ICT-215078 DIPLECS, by the Czech Science Foundation Project 102/07/1317, and by the Grant Agency of the CTU Prague under Project SGS10/069/OHK3/1T/13.

#### REFERENCES

[1] J. Matas, O. Chum, M. Urban, and T. Pajdla, "Robust wide baseline stereo from maximally stable extremal regions," in *British Machine Vision Conference*, vol. 1, 2002, pp. 384–393.

image 1 $\leftrightarrow$ image 2 from Graffiti					
method	all			best 500 tc	
	tc	inl	inl/tc	inl	inl/tc
plain	622	425	68.3%	418	83.6%
smooth	678	516	76.1%	458	91.6%
refine	921	<b>702</b>	<b>76.2%</b>	<b>486</b>	<b>97.2%</b>

image 1 $\leftrightarrow$ image 3 from Graffiti					
method	all			best 500 tc	
	tc	inl	inl/tc	inl	inl/tc
plain	500	284	56.8%	284	56.8%
smooth	566	337	<b>59.5%</b>	337	67.4%
refine	801	<b>470</b>	58.7%	<b>409</b>	<b>81.8%</b>

Table I  
MATCHING RESULTS FOR GEOMETRIC HASHING WITH LAFs, TC - NUMBER OF TENTATIVE CORRESPONDENCES, INL - NUMBER OF INLIERS.

- [2] 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*, vol. 65, no. 1-2, pp. 43–72, 2005.
- [3] Š. Obdržálek, "Object recognition using local affine frames," PhD Thesis, 2007.
- [4] F. Mokhtarian and A. K. Mackworth, "A theory of multiscale, curvature-based shape representation for planar curves," *Pattern Analysis and Machine Intelligence*, vol. 14, no. 8, pp. 789–805, 1992.
- [5] Š. Obdržálek and J. Matas, "Object recognition using local affine frames on distinguished regions," in *Proceedings of the British Machine Vision Conference*, P. L. Rosin and D. Marshall, Eds., vol. 1. London, UK: BMVA, September 2002, pp. 113–122.
- [6] H. Shao, T. Svoboda, and L. Van Gool, "Zubud — zurich buildings database for image based recognition," Computer Vision Lab, Swiss Federal Institute of Technology, Switzerland, Tech. Rep., 2003.
- [7] L. Roberts, "Machine perception of three dimensional solids," MIT Lincoln Lab., Tech. Rep., May 1963.
- [8] A. Mikulík, "Methods for precise local affine frame constructions on msers," Master's thesis, Charles University in Prague, Faculty of Mathematics and Physics, 2009.
- [9] R. Hartley and A. Zisserman, *Multiple view geometry in computer vision*, 2nd ed. Cambridge: Cambridge University, 2003.
- [10] O. Chum and J. Matas, "Geometric hashing with local affine frames," in *Computer Vision and Pattern Recognition*, June 2006, pp. 879–884.
- [11] J. Matas and O. Chum, "Randomized ransac with sequential probability ratio test," in *International Conference on Computer Vision*, 2005.