Two-view Matching with View Synthesis Revisited

Dmytro Mishkin, Michal Perdoch, Jiri Matas
Center for Machine Perception, Faculty of Electrical Engineering, Czech Technical University in Prague
ducha.aiki@gmail.com, {perdom1,matas}@cmp.felk.cvut.cz

***Abstract*—Wide-baseline matching focussing on problems with extreme viewpoint change is considered. We introduce the use of view synthesis with affine-covariant detectors to solve such problems and show that matching with the Hessian-Affine or MSER detectors outperforms the state-of-the-art ASIFT [19].**

To minimise the loss of speed caused by view synthesis, we propose the Matching On Demand with view Synthesis algorithm (MODS) that uses progressively more synthesized images and more (time-consuming) detectors until reliable estimation of geometry is possible. We show experimentally that the MODS algorithm solves problems beyond the state-of-the-art and yet is comparable in speed to standard wide-baseline matchers on simpler problems.

Minor contributions include an improved method for tentative correspondence selection, applicable both with and without view synthesis and a view synthesis setup greatly improving MSER robustness to blur and scale change that increase its running time by 10% only.

***Keywords*—feature extraction, image matching, view synthesis.**

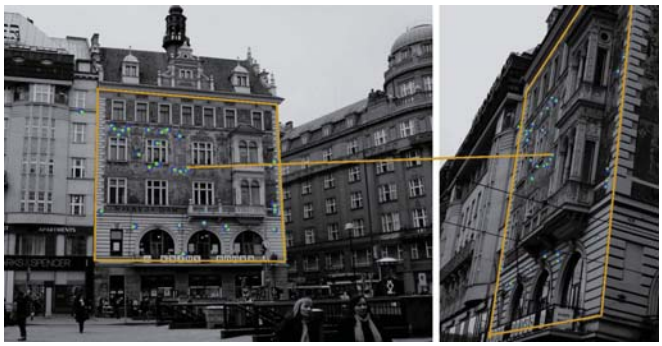


Fig. 1. Homography estimation with extreme viewpoint change. The proposed MODS algorithm produces 32 matches, 25 are correct. The state-of-the-art ASIFT [19] outputs 41 matches, 3 are correct. Blue dots: centers of detected regions. Green dots: reprojected centers of corresponding regions showing good alignment.

I. Introduction

The standard method for wide baseline matching involves detection of local features, calculation of descriptors, generation of tentative correspondences and their geometric verification using the homography or epipolar constraint. It is well known [7], [8], [17] that performance of the pipeline decreases in the presence of viewpoint and scale changes, blur, compression artefacts, etc. Lepetit and Fua [12] showed that matching robustness is improved by synthesis of additional views given a single, fronto-parallel view of an object. Morel and Yu [19] combined viewpoint synthesis with the similarity-covariant Difference-of-Gaussians detector (DoG) and SIFT matching [14]. The resulting image matching method, called ASIFT, successfully matched challenging image pairs with significantly different viewing angles.

We develop the idea of view synthesis for wide baseline matching and propose a number of novelties that improve several stages of the matching pipeline. Some of the improvements are also applicable to two-view matching without synthesis. The proposed MODS wide-baseline matcher[1] outperforms ASIFT in terms of speed, the number and percentage of correct matches generated as well as in the precision of the estimated geometry. Performance was tested mainly on image pairs with extreme viewpoint changes, but viewpoint synthesis also improves matching results in the presence of phenomena like blur, occlusion and scale change. The following contributions are made: first, we show that the seemingly counter-intuitive synthesis of affine views for "affine-covariant" detectors greatly improves their performance in wide baseline matching. With suitable detector-specific configurations of synthesized viewpoints, found through extensive experimentation, both the Hessian-Affine [16] and MSER [15] detectors clearly outperform DoG [14].

Second, we generalize the "first-to-second-closest SIFT distance ratio" criterion for the selection of tentative correspondences. Depending on the image, the new criterion gives 5-20% more true matches than the standard at no extra computation cost. The proposed criterion improves even matching performance without synthesis, especially in images with local symmetries.

Third, we propose an adaptive algorithm for matching very challenging image pairs which follows the "do only as much as needed" principle. The MODS algorithm (Matching On Demand with view Synthesis) uses progressively more detector types and more synthesized images until enough correspondences for reliable estimation of two-view geometry are found. MODS is fast on easy image pairs without compromising performance on the hardest problems.

A. Related work

The use of view synthesis for image matching is a recent development and the literature is limited and includes mainly modifications of the ASIFT algorithm. Liu *et al*. [13] synthesised perspective warps rather than affine. Pang *et al*. [21] replaced SIFT by SURF [3] in the ASIFT algorithm to reduce the computation time. Sadek *et al*. [23] present a new affine covariant descriptor based on SIFT which can be used with or

978-1-4799-0883-7/13/$31.00 ©2013 IEEE
[1]Available at http://cmp.felk.cvut.cz/wbs/index.html

without view synthesis. Detection of the MSERs on the scale space pyramid was proposed by Forssén and Lowe [9].

II. Matching with On Demand View Synthesis

The iterative MODS algorithm (see Alg. 1) repeats a sequences of two-view matching procedures, until a required minimum number of geometrically verified correspondences is found. In each iteration, a different detector is used and a different set of views generated. The adopted sequence is an outcome of extensive experimentation with the objective of solving the most challenging problems while keeping speed comparable to standard single-detector wide-baseline matchers for simple problems. For instance, the first iteration of the MODS algorithm runs the MSER detector with only a very coarse scale space pyramid which is 10% slower than standard MSER. Subsequent iterations run complementary detectors with a higher number of synthesized views. The rest of the section describes the steps employed in the iterations of the MODS algorithm.

Algorithm 1 MODS: Matching with On-Demand view Synthesis

Input: I_1, I_2 – two images; θ_m – minimum required number of matches; $S_{\max}$ – maximum number of iterations.
Output: Fundamental of homography matrix F or H; list of corresponding points.
Variables: N_{matches} – detected correspondences, Iter – currect iteration.

```
while (N_matches < θ_m) and (Iter < S_max) do
  for I_1 and I_2 separately do
    1 Generate synthetic views according to the
      scale-tilt-rotation-detector setup for the Iter.
    2 Detect and describe local features.
    3 Reproject local features to original image.
      Add described features to general list.
  end for
  4 Generate tentative correspondences
    using the first geom. inconsistent rule.
  5 Filter duplicate matches.
  6 Geometrically verify tentative correspondences
    while estimating F or H.
end while
```

A. Synthetic views generation

It is well known that a homography H can be approximated by an affine transformation A at a point using the first order Taylor expansion. Further, an affine transformation can be uniquely decomposed by SVD into a rotation, skew, scale and rotation around the optical axis [10]. Morel and Yu [19] proposed to decompose the affine transformation A as

$$\begin{aligned} A &= H_\lambda R_1(\psi) T_t R_2(\phi) = \\ &= \lambda \begin{pmatrix} \cos\psi & -\sin\psi \\ \sin\psi & \cos\psi \end{pmatrix} \begin{pmatrix} t & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} \cos\phi & -\sin\phi \\ \sin\phi & \cos\phi \end{pmatrix} \end{aligned} \quad (1)$$

where $\lambda > 0$, R_1 and R_2 are rotations, and T_t is a diagonal matrix with $t > 1$. Parameter t is called the absolute tilt, $\phi \in \langle 0, \pi)$ is the optical axis longitude and $\psi \in \langle 0, 2\pi)$ is the rotation of the camera around the optical axis. Each synthesised view is parametrised by the tilt, longitude and optionally the scale and represents a sample of the view-sphere resp. view-volume around the original image.

The view synthesis proceeds in the following steps: at first, scale synthesis is performed by building a Gaussian scale-space with Gaussian $\sigma = \sigma_{\text{base}} \cdot S$ and downsampling factor S ($S < 1$). Second, each image in the scale-space is in-plane rotated by longitude ϕ with step $\Delta\phi = \Delta\phi_{\text{base}}/t$. In the third step, all rotated images are convolved with a Gaussian filter with $\sigma = \sigma_{\text{base}}$ along vertical direction and $\sigma = t \cdot \sigma_{\text{base}}$ along horizontal direction to eliminate aliasing in the final tilting step. The tilt is applied by shrinking the image along the horizontal direction by factor t. The parameters of the synthesis are: the set of scales $\{\text{S}\}$, $\Delta\phi_{\text{base}}$ – the step of longitude samples at tilt $t = 1$, and a set of simulated tilts $\{\text{t}\}$. The details of view synthesis parameter tuning for each detector are presented in technical report [18].

B. Local feature detection and description

The goal of the view synthesis procedure is to provide detectors with a sufficiently similar subset of all artificial views on the view-sphere that allows matching. For affine-covariant detectors, unlike the similarity-covariant DoG of ASIFT, the number of necessary view samples is significantly decreased while the performance for the most difficult image pairs gets improved. Moreover, it is known that different detectors are suitable for different types of images [17] and that some detectors are complementary in the feature points they detect [1]. Our experiments show (*c.f.* Section III) that combining detectors improves the overall robustness and speed of the matching procedure.

MODS uses the state-of-the-art affine covariant detectors MSER and Hessian-Affine. The normalised patches are described by the recent modification of SIFT [14] – the RootSIFT [2]. The local feature frames computed on the synthesised views are backprojected to the coordinate system of the original image by a known affine matrix A and associated with the descriptor and the originating synthetic view.

C. Tentative correspondence generation

Different strategies for computation of the tentative correspondences in wide-baseline matching have been proposed. The standard method for matching SIFT(-like) descriptors is based on the distance ratio of the closest to the second closest descriptors in the other image [14]. Performance of this test in general very efficient method degrades when multiple observations of the same feature are present. In this case, the similar descriptors will lead to the first to second SIFT ratio to be close to 1 and the correspondences will "annihilate" each other, despite the fact they all represent the same geometric constraints and are therefore not mutually contradictory (see Figure 2). The problem of multiple detections is amplified in the matching by view synthesis since covariantly detected local features have often a response in multiple synthetic views. We propose to use, instead of comparing the first to the second closest descriptor distance, the distance of the first descriptor and the closest descriptor that is geometrically inconsistent with the first one (denoted 1st inc. in the following). We call descriptors in one image *geometrically inconsistent* if the Euclidean distance between centers of the regions is ≥ 10 pixels. The difference of the first-to-second closest ratio strategy and the closest-to-1st inc. strategy is illustrated in Figure 2.

The kd-tree algorithm from FLANN library [20] effectively finds the N-closest descriptors in the other image. The distance ratio thresholds of the closest to 1st inc. were experimentally selected based on the CDFs of matching and non-matching

Fig. 2. Comparison of the proposed *first to 1st inc. ratio* matching strategy and the standard *first to second closest ratio* matching strategy. Red regions are the second closest descriptors, yellow regions correspond to the closest geometrically inconsistent descriptors, green are the true corresponding regions. Upper pair – rotationally symmetric DoG regions, lower pair – affine covariant MSER regions.

descriptors (see [18]). We recommend to use the same values for SIFT and RootSIFT descriptors, but different thresholds for the different local feature detectors: $R_{\text{MSER}} = 0.85$, $R_{\text{DoG}} = 0.85$ and $R_{\text{HA}} = 0.8$.

D. Duplicate filtering

The redetection of covariant features in synthetic views results in duplicates in tentative correspondences. The duplicate filtering is an optional step and prunes correspondences with close spatial distance of local features in both images. The number of pruned correspondences can be however used later for evaluating the quality (probability of being correct) in PROSAC-like [4] geometric verification.

E. Geometric verification

The LO-RANSAC [11] algorithm searches for the maximal set of geometrically consistent tentative correspondences. The model of the transformation is set either to homography or epipolar geometry, or automatically determined by a Degen-SAC [5] procedure.

III. Experiments

A. 1st geometrically inconsistent vs. 2nd nearest neighbour correspondence selection strategy

The *first to first geometrically inconsistent* strategy was evaluated on 50 image pairs of the publicly available datasets [17] and [6]. The cumulative distributions of the number of correct tentative correspondences as functions of the descriptor distance ratio are used for comparison. The new matching strategy improves the performance by up to 5% for the matching without view synthesis and up to 30% (see Figure 3) for matching with view synthesis at almost no additional computational costs.

B. View synthesis for affine covariant detectors

The view synthesis parameters – tilt $\{t\}$ sampling and longitude step $\Delta\phi_{\text{base}}$ – were explored in the following synthetic

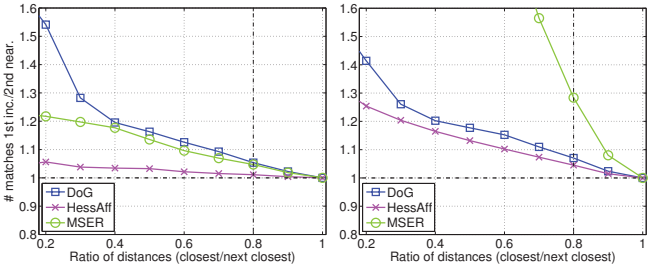


Fig. 3. The ratio of the number of correct matches obtained by the 1st inconsistent and 2nd nearest method, without (left) and with (right) view synthesis. The black dashed line denotes the widely used distance ratio threshold = 0.8.

experiment. For each of 100 random images from Oxford Building Dataset [22], a set of simulated views with latitude angle $\theta = (0, 20, 40, 60, 65, 70, 75, 80, 85)^\circ$, corresponding to tilt series $t = (1.00, 1.06, 1.30, 2.00, 2.36, 2.92, 3.86, 5.75, 11.47)$[2] was generated. The ground truth affine matrix A was computed for each synthetic view using equation (1). The original and synthesised images were matched using described algorithm with single iteration.

The various configurations of the view synthesis were tested and results for the selected configurations are shown in Figure 4. Note that the view synthesis significantly increases the matching performance, however after reaching some density of the view-sphere sampling additional views does not bring more correspondences. MSER and Hessian-Affine need sparser view-sphere sampling than DoG. Results for all tested configurations are in technical report [18].

C. Results on the Extreme Viewpoint Dataset

We introduce a two-view matching evaluation dataset[3] with extreme viewpoint changes, see Table I. The dataset includes image pairs from publicly available datasets: ADAM and MAG [19], GRAF [17] and THERE [6]. The ground truth homography matrices were estimated by LO-RANSAC using correspondences from all three detectors in view synthesis configuration $\{t\} = \{1; \sqrt{2}; 2; 2\sqrt{2}; 4; 4\sqrt{2}; 8\}$, $\Delta\phi = 72^\circ/t$. The number of inliers for each image pair was ≥ 50 and the homographies were manually inspected. For the image pairs GRAF and THERE precise homographies are provided by Cordes *et al.* [6]. Transition tilts τ were computed using equation (1) with SVD decomposition of the linearised homography at center of the first image of the pair (see Table I).

The configurations evaluated are specified in Table II. For comparison, ASIFT[4] results are added. Computations were performed on Intel i5 CPU @ 2.6GHz with 4Gb RAM; results for computation on one core are provided. Based on results of the different configuration, we have chosen the following configuration for MODS w.r.t increasing computation time and performance of the configurations – see Table III. Please note that only views complementary to the previous iterations are synthesised.

[2] assuming that the original image is in the fronto-parallel view

[3] Available at http://cmp.felk.cvut.cz/wbs/index.html

[4] Reference code from http://demo.ipol.im/demo/my_affine_sift

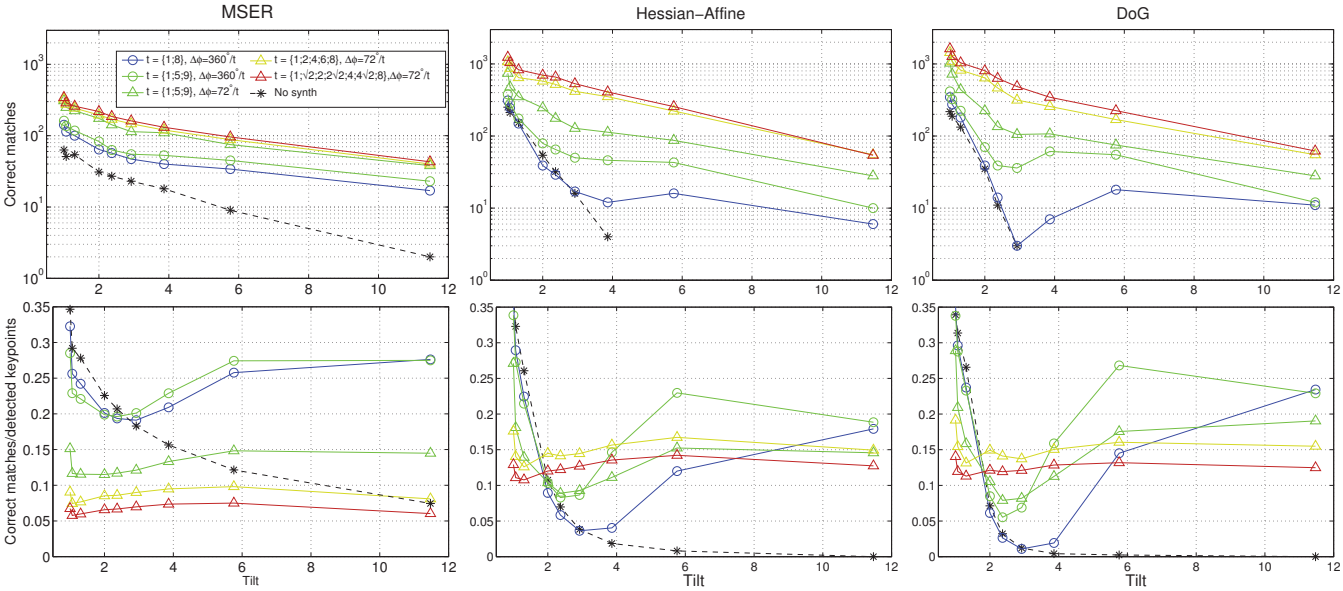


Fig. 4. Comparison of view synthesis configurations on the synthetic dataset. First row: the number of correct SIFT matches a robust minimum (value 4% quantile) over 100 random images from [22]). Second row: the ratio of the number of correct matches to the number of detected regions; the mean over 100 random images. Only selected configurations are shown, full version is in [18].

TABLE II. VIEW SYNTHESIS CONFIGURATIONS BASED ON THE ANALYSIS OF THE ALGORITHM ON THE SYNTHETIC DATASET

	Configurations	
Detector	SPARSE	DENSE
MSER	$\{S\} = \{1; 0.25; 0.125\}$, $\{t\} = \{1; 5; 9\}$, $\Delta\phi = 360°/t$	$\{S\} = \{1; 0.25; 0.125\}$, $\{t\} = \{1; 2; 4; 6; 8\}$, $\Delta\phi = 72°/t$
HessAff	$\{S\} = \{1\}$, $\{t\} = \{1; \sqrt{2}; 2; 2\sqrt{2}; 4; 4\sqrt{2}; 8\}$, $\Delta\phi = 360°/t$	$\{S\} = \{1\}$, $\{t\} = \{1; 2; 4; 6; 8\}$, $\Delta\phi = 72°/t$
DoG	$\{S\} = \{1\}$, $\{t\} = \{1; 2; 4; 6; 8\}$, $\Delta\phi = 120°/t$	$\{S\} = \{1\}$, $\{t\} = \{1; \sqrt{2}; 2; 2\sqrt{2}; 4; 4\sqrt{2}; 8\}$, $\Delta\phi = 72°/t$

TABLE III. CONFIGURATIONS FOR MODS STEPS

Iter.	Setup
1	MSER,$\{S\} = \{1; 0.25; 0.125\}$, $\{t\} = \{1\}$, $\Delta\phi = 360°/t$
2	MSER,$\{S\} = \{1; 0.25; 0.125\}$, $\{t\} = \{1; 5; 9\}$, $\Delta\phi = 360°/t$
3	HessAff, $\{S\} = \{1\}$, $\{t\} = \{1; \sqrt{2}; 2; 2\sqrt{2}; 4; 4\sqrt{2}; 8\}$, $\Delta\phi = 360°/t$
4	HessAff , $\{S\} = \{1\}$, $\{t\} = \{1; 2; 4; 6; 8\}$, $\Delta\phi = 72°/t$

The MODS algorithm allows to set the minimum desired number of inliers as a stopping criterion. The recommended value – 15 inliers to the homography, have a very low probability to be a random result, but are few enough to show the time gain from the algorithm. To maximize the number of inliers for each of the detectors, we recommend to use DENSE configuration as a single step. Figure 5 and Table IV compare the different view synthesis configurations and the "affine-covariant" detectors – they generate more correct matches in a shorter time than the DoG detector. The DoG based matching and ASIFT matching cannot solve 3 resp. 9 out of the 15 image pairs. The ASIFT algorithm generates a lower number of correct inliers and works slower than our DoG DENSE configuration (which has the same tilt-rotation set). The main causes are elimination of "one-to-many", including correct, correspondences, the inferiority of the standard 2nd closest ratio and a simple bruteforce algorithm of matching used in ASIFT.

No single detector solved all image pairs. The Hessian-Affine with DENSE configuration successfully solved 14 out of 15 image pairs and outperformed other detectors and configurations in the number of inliers, however, at the expense of the highest computational cost. MSER with no synthesis and in the SPARSE configuration is the fastest and could solve 10 out of 15 image pairs. The MODS algorithm solves all image pairs and saves computational time on processing of the easy pairs at the cost of a small matching overhead on the hard cases. Also, MODS is the fastest algorithm in 7 cases, and in another 2 cases it is just $\sim 10\%$ slower than the fastest configuration. The difference results of MODS step 2 and MSER SPARSE are caused by randomization in RANSAC and kd-tree building.

Fig. 6 shows the breakdown of the computational time. SIFT description with the dominant orientation estimation take 50% of the time. Note that the whole process is almost linear in the area of the synthesised views. The only super-linear part, matching, takes only 10% of the time.

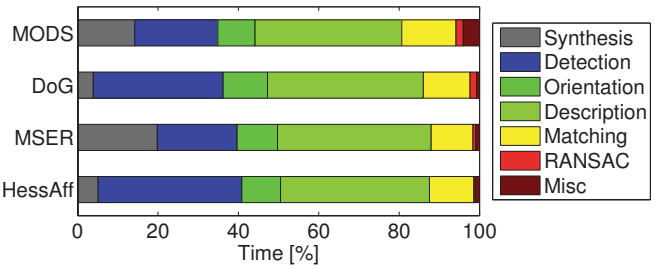
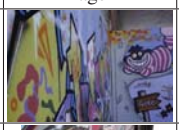
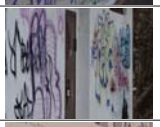
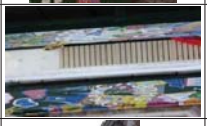


Fig. 6. Percentage of time spent in the main stages of the matching with view synthesis process on a single core, DENSE configuration. SIFT description, i.e. the dominant gradient estimation and the descriptor computation is the most time-consuming part.

TABLE I. THE EXTREME VIEW DATASET – EVD. IMAGE SOURCES: C – CORDES *et al.* [6], OX – MIKOLAJCZYK *et al.* [17], M – MOREL AND YU [19].

#	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Name	THERE	GRAF	ADAM	MAG	GRAND	PKK	FACE	GIRL	SHOP	DUM	INDEX	CAFE	FOX	CAT	VIN
Source	C	Ox	M	M	EVD	EVD	EVD	EVD	EVD	EVD	EVD	EVD	EVD	EVD	EVD
τ – transitional tilt	6.3	3.6	4.8	20	2.9	7.1	6.9	8.0	9.1	6.9	8.5	11.9	22.5	47	49.8

#	Image 1	Image 2	#	Image 1	Image 2	#	Image 1	Image 2
1			6			11		
2			7			12		
3			8			13		
4			9			14		
5			10			15		

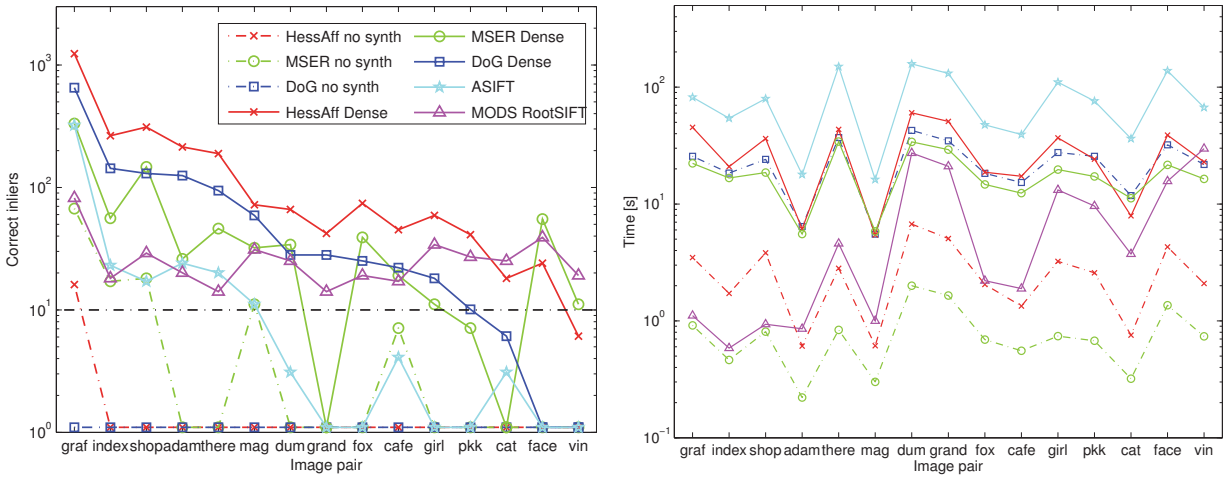


Fig. 5. Performance of the selected view synthesis configurations defined in Table II. MODS set to find ≥ 15 inliers. Left – the number of correct RANSAC inliers. The black dashed line marks the level of 10 correct inlier – a minimum for a reliable estimate of two-view geometry. Right – runtime (1 core).

D. MSER vs. blur and scale change

We have tested performance of recommended scale synthesis configuration for MSER on the image pairs most distorted by blur and scale change from the Oxford [17] dataset. To allow comparison with [17], the standard SIFT was used instead of RootSIFT in this experiment. Note that the results are not fully compatible as we use NN-distance ratio matching threshold = 0.8 (In [17] no ratio threshold has been used, so the absolute number of the matches differs a lot. But relative ratio between detectors performance remains the same). We have also performed the duplicate filtering procedure, which reduces the number of correspondences (*c.f.* Section II).

Figure 7 shows that scale synthesis with 1st geom. inconsistent rule improves MSER performance by 60% to 1000%, solving the most common MSER problems – sensitivity to blur and scale change. The quality of tentative correspondences also increases with the proposed scale synthesis configuration (Figure 7, right). Table V shows the computation time.

TABLE V. MSER MATCHER RUNTIME ON OXFORD [17] DATASET

scale synthesis setup	time [s]
$\{S\} = \{1\}$	56.6
$\{S\} = \{1; 0.25; 0.125\}$	61.5

IV. CONCLUSIONS

We have introduced view synthesis to two-view wide-baseline matching with affine-covariant detectors and shown

TABLE IV. A COMPARISON OF DIFFERENT VIEW SYNTHESIS AND DETECTOR CONFIGURATIONS (WITH ROOTSIFT). BEST RESULTS ARE HIGHLIGHTED BY A GREY BACKGROUND. MODS SET TO FIND ≥ 15 INLIERS. RESULTS WITH LESS THAN 8 CORRECT INLIERS ARE IN RED.

Image	Correct inliers						Time, 1 core [s]						Correct inliers/sec					
	MODS, $\theta_m = 15$	ASIFT	MSER SPARSE	HessAff SPARSE	HessAff DENSE	DoG DENSE	MODS, $\theta_m = 15$	ASIFT	MSER SPARSE	HessAff SPARSE	HessAff DENSE	DoG DENSE	MODS, $\theta_m = 15$	ASIFT	MSER SPARSE	HessAff SPARSE	HessAff DENSE	DoG DENSE
graf	82	322	165	375	1235	653	1.0	81.8	3.0	11.0	45.2	25.5	83.9	3.9	55	34.1	27.3	25.6
index	18	23	24	34	264	143	0.5	54.1	2.2	5.4	20.8	18.3	38.1	0.4	11.1	6.3	12.7	7.8
shop	29	17	73	133	311	130	0.8	79.5	2.5	10.1	36.2	24	35.2	0.2	28.7	13.2	8.6	5.4
adam	20	24	18	86	214	125	0.8	17.8	0.7	1.6	6.0	6.3	26.7	1.3	24.3	54.1	35.6	19.8
there	14	20	12	49	189	94	4.5	150.0	4.5	10.1	43.4	36.9	3.1	0.1	2.7	4.9	4.4	2.5
mag	31	11	28	54	72	59	0.8	16.1	0.8	1.6	5.3	5.4	37.3	0.7	34.4	33.5	13.5	10.9
dum	25	3	0	10	66	28	29.4	158.0	4.8	20.1	60.2	42.5	0.9	0.0	0.0	0.5	1.1	0.7
grand	14	0	9	0	42	28	21.9	131.0	4.2	14.8	50.8	34.6	0.6	0.0	2.1	0.0	0.8	0.8
fox	19	0	19	22	74	25	2.1	47.4	2.1	5.8	18.6	18.2	9.0	0.0	9.3	3.8	4	1.4
cafe	17	4	14	0	45	22	1.8	39.2	1.7	4.5	17.2	15.2	9.3	0.1	8.2	0.0	2.6	1.4
girl	34	0	0	14	59	18	13.1	110.0	2.7	10.0	36.7	27.5	2.6	0.0	0.0	1.4	1.6	0.7
pkk	27	0	6	12	41	10	9.5	75.9	2.4	6.8	24.1	25.5	2.8	0.0	2.5	1.8	1.7	0.4
cat	25	3	0	21	18	6	3.9	36.2	1.4	2.2	7.8	11.7	6.3	0.1	0.0	9.6	2.3	0.5
face	39	0	9	17	24	0	15.6	138.0	3.4	11.3	38.8	32.0	2.5	0.0	2.7	1.5	0.6	0.0
vin	19	0	0	0	6	0	30.3	66.9	2.3	6.3	22.8	21.7	0.6	0.0	0.0	0.0	0.3	0.0

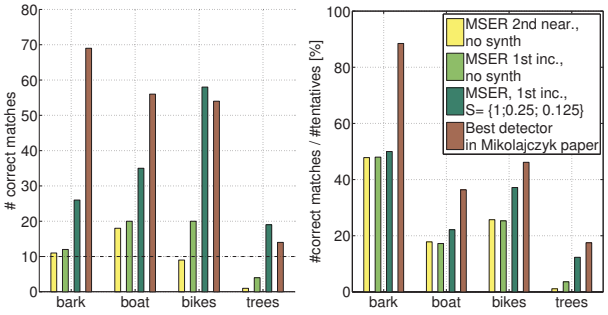


Fig. 7. MSER performance with and w/o scale synthesis on the most distorted pairs (1-6) with scale change and blur from [17]. Left – the number of correct SIFT matches. Right – the proportion of correct matches within tentative correspondences. The best detectors from [17]: BARK, BOAT, TREES – Hessian-Affine, BIKES – IBR are shown for comparison.

that matching with the Hessian-Affine or MSER detectors outperforms the state-of-the-art ASIFT.

To address the robustness vs. speed trade-off, we have proposed the Matching On Demand with view Synthesis algorithm (MODS) that uses progressively more synthesized images and more (time-consuming) detectors until a reliable estimate of geometry is obtained. We show experimentally that the MODS algorithm solves matching problems beyond the state-of-the-art and yet is comparable in speed to standard wide-baseline matchers on simpler problems.

Minor contributions include an improved method for tentative correspondence selection, applicable both with and without view synthesis. A modification of the standard first to second nearest SIFT distance rule increases the number of correct matches by 5-20% at no additional computational cost. Finally, we found a simple view synthesis set up costing less than 10% of time that greatly improves MSER robustness to blur and scale change.

ACKNOWLEDGMENT

The authors were supported by EC project FP7-ICT-270138 DARWIN the Technology Agency of the Czech Republic project TE01020415 V3C and the MSMT grant LL1303 ERC-CZ.

REFERENCES

[1] H. Aanaes, A. Dahl, and K. Steenstrup Pedersen. Interesting interest points. *IJCV*, 97(1):18–35, 2012.
[2] R. Arandjelović and A. Zisserman. Three things everyone should know to improve object retrieval. In *CVPR*, 2012.
[3] H. Bay, T. Tuytelaars, and L. V. Gool. SURF: Speeded up robust features. In *ECCV*, 2006.
[4] O. Chum and J. Matas. Matching with PROSAC – progressive sample consensus. In *CVPR*, 2005.
[5] O. Chum, T. Werner and J. Matas. Two-view geometry estimation unaffected by a dominant plane. In *CVPR*, 2005.
[6] K. Cordes, B. Rosenhahn, and J. Ostermann. Increasing the accuracy of feature evaluation benchmarks using differential evolution. In *SSCI-Symposium on Differential Evolution*, 2011.
[7] A. L. Dahl, H. Aanaes, and K. S. Pedersen. Finding the best feature detector-descriptor combination. *3DIMPVT*, 2011.
[8] F. Fraundorfer and H. Bischof. A novel performance evaluation method of local detectors on non-planar scenes. In CVPR'05 Workshops, 2005.
[9] P.-E. Forssén and D. Lowe. Shape Descriptors for Maximally Stable Extremal Regions. In ICCV, 2007.
[10] R. I. Hartley and A. Zisserman. *Multiple View Geometry in Computer Vision*, 2004.
[11] K. Lebeda, J. Matas, and O. Chum. Fixing the Locally Optimized RANSAC. In *BMVC*, 2012.
[12] V. Lepetit and P. Fua. Keypoint recognition using randomized trees. *PAMI*, 28(9):1465–1479,2006.
[13] W. Liu, Y. Wang, J. Chen, J. Guo, and Y. Lu. A completely affine invariant image-matching method based on perspective projection. *MVA*, 23(2):231–242, 2012.
[14] D. Lowe. Distinctive image features from scale-invariant keypoints. *IJCV*, 60(2):91–110, 2004.
[15] J. Matas, O. Chum, M. Urban, and T. Pajdla. Robust wide-baseline stereo from maximally stable extremal regions. In *BMVC*, 2002.
[16] K. Mikolajczyk and C. Schmid. Scale & affine invariant interest point detectors. *IJCV*, 60(1):63–86,2004.
[17] K. Mikolajczyk, T. Tuytelaars, C. Schmid, A. Zisserman, J. Matas, F. Schaffalitzky, T. Kadir, and L. V. Gool. A comparison of affine region detectors. *IJCV*, 65(1-2):43–72, 2005.
[18] D. Mishkin, M. Perdoch, and J. Matas. Two-view matching with view synthesis revisited. Tech. Rep. *CoRR*, abs/1306.3855, 2013.
[19] J.-M. Morel and G. Yu. ASIFT: A new framework for fully affine invariant image comparison. *SIIMS*, 2(2):438–469, 2009.
[20] M. Muja and D. G. Lowe. Fast approximate nearest neighbors with automatic algorithm configuration. In *VISSAPP'09)*, 2009.
[21] Y. Pang, W. Li, Y. Yuan, and J. Pan. Fully affine invariant SURF for image matching. *Neurocomputing*, 85(0):6–10, 2012.
[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] R. Sadek, C. Constantinopoulos, E. Meinhardt, C. Ballester, and V. Caselles. On affine invariant descriptors related to SIFT. *SIIMS*, 5(2):652–687,2012.