Localization Accuracy of Region Detectors

Andreas Haja, Bernd Jähne
Interdisciplinary Center for Scientific Computing
University of Heidelberg, Speyerer Straße 4-6
69115 Heidelberg, Germany
Firstname.Lastname@iwr.uni-heidelberg.de

Steffen Abraham
Robert Bosch GmbH, Corporate Research
Robert-Bosch-Straße 200
31139 Hildesheim, Germany
Steffen.Abraham@de.bosch.com

Abstract

In this paper, a comparison of five state of the art region detectors is presented with regard to localization accuracy in position and region shape. Based on carefully estimated ground truth homographies, correspondences between frames are assigned using geometrical region overlap. Significant differences between detectors exist, depending on the type of images. Also, it is shown that localization accuracy linearly depends on region scale for some detectors, which may thus be used as a pre-selection criterion for the removal of error-prone regions. The presented results serve as a supplement to existing comparative studies, and can be used to facilitate the selection of an appropriate detector for a specific target application. When descriptor distance is used as assignment criterion instead of region overlap, a different set of correspondences results with lower accuracy. Set differences (and thus localization accuracy) are directly related to the density of regions in a local neighborhood. Based on the latter, a novel measure for the identification of error-prone regions - shape uniqueness - is introduced. In contrast to existing methods that are based on the descriptor distance of region correspondences, the new measure is pre-computed on each image individually. Thus, the complexity of the subsequent matching task can be significantly reduced.

1. Introduction

In this work, five popular region detectors are compared with regard to localization accuracy in position and region shape: the edge-based (EBR) and intensity-based (IBR) region detectors by Tuytelaars and Van Gool[10], the maximally stable extremal regions detector (MSER) by Matas *et al*. [5], and both Harris- and Hessian-affine detectors (HARAFF & HESAFF) introduced by Mikolajczyck and Schmid[6]. For all experiments, the original implementations[1] of the authors with their respective default parameters are used (including the SIFT-descriptor).

A thorough analysis of state of the art region detectors has been published by Mikolajczyck et al. in [7]. There, detector performance is evaluated with regard to feature repeatability under several image transformations. The underlying idea is that under a geometric or photometric transformation between two images, the set of regions detected in the first frame should correspond to the set of transformed regions in the second frame. The authors introduce the *repeatability score* as a measure for detector performance, defined as the ratio between the number of region correspondences and the smaller number of regions in a pair of images. A good detector should exhibit both a large number of correspondences and a high repeatability score. Moreels and Perona [8] additionally evaluated the performance of region detectors and descriptors, but based on 3D-objects instead of planar objects.

Until now, there exists only little information on how accurately regions are located. Therefore, detectors are compared within this work with regard to their localization accuracy in position and region shape. If reliable information on the latter were available, depth changes of moving objects could be observed from the relative scale difference of region correspondences, *e.g*. in a tracking application. To the best knowledge of the authors, such a comparative study has not yet been published in the literature. The presented results serve as a supplement to existing comparative studies, and can be used to facilitate the selection of appropriate region detectors for a specific target application.

Firstly, based on carefully estimated ground truth homographies, correspondences between frames are assigned using geometrical region overlap. It is demonstrated, that significant differences between detectors exist, depending on the type of images. Also, it is shown that localization accuracy linearly depends on region scale for some detectors, which may thus be used as a pre-selection criterion for the removal of error-prone regions. Secondly, descriptor distance is used as assignment criterion instead of region overlap (based on the well-known SIFT-method [4]). The result-

[1] http://www.robots.ox.ac.uk/ vgg/research/affine

978-1-4244-2243-2/08/$25.00 ©2008 IEEE

ing set of correspondences differs from the overlap-based set and generally exhibits a lower accuracy. It is shown that set differences and accuracy are directly related to region density. Based on the latter, a novel measure for the identification of error-prone regions - shape uniqueness - is introduced. In contrast to existing methods that are based on the descriptor distance of region correspondences, the new measure is pre-computed on each image individually. By removing regions with low shape uniqueness, the complexity of the matching task can be significantly reduced, while the resulting set of final correspondences stays similar.

This paper is organized as follows: In section 2, the evaluated image sequences are shown. In section 3, a method for the selection of region correspondences is introduced, along with a first qualitative assessment of region accuracy. Further, a detailed evaluation of errors in both position and shape is presented in section 4. Finally, section 5 gives a concluding summary and an outlook on future work.

2. The Image Data Set

Figure 1 shows a reduced version of the image sets analyzed within this work. Each sequence contains 6 roughly equal-sized gray scale images ($\approx 800 \times 640$ pel). The sequences *boat*, *wall* and *graffiti* were taken from the dataset provided by Mikolajczyck *et al.*[1]. All others were chosen from the database made available by Moreels and Alatorre[2].

The 6 sequences are categorized into two disjoint groups: *graffiti*, *carton* and *frame* mainly contain homogeneous regions with distinctive edges (*structured*), whereas *dvd*, *boat* and *wall* show repeating patterns and natural textures (*textured*). All experiments in sections 3 and 4.1 are conducted separately on both groups. Every image shows mainly planar objects so that carefully estimated homographies may be used to determine region correspondences. If a scene is not planar (as with *boat*), the camera position has been fixed during image acquisition. The homographies for *boat*, *wall* and *graffiti* have been made available online[1], for all other sequences we estimated them ourselves according to the method described in [7]. The *root mean square* error is less than 1 pel for each image pair. In order to eliminate the influence of color information, all images have been converted to gray scale.

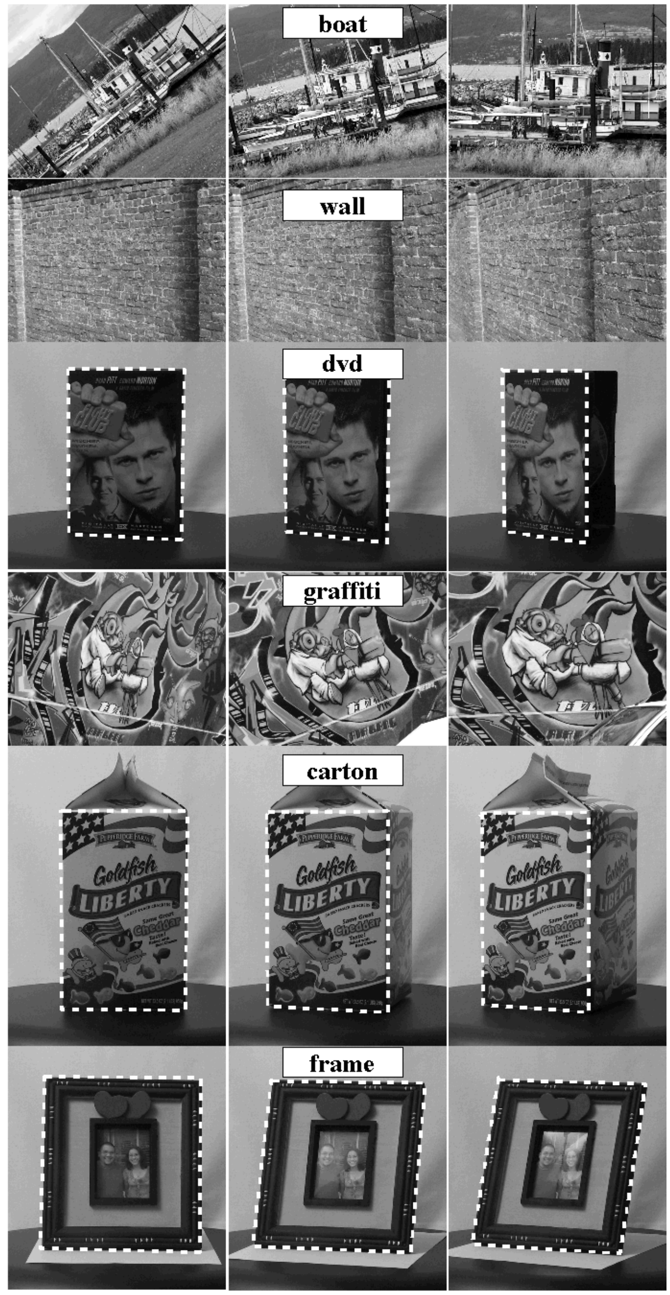

Figure 1. A reduced version of the image set: The top 3 sequences are categorized as *textured* with repeating patterns and natural textures, whereas the bottom 3 sequences are classified as *structured*, with distinctive edges and homogeneous regions.

3. Region Correspondences

In this section, a method for the selection of correspondences is introduced, based on the two-dimensional area overlap of candidate regions. Further, the susceptibility of all detectors to changes in the maximally permissible overlap is assessed, enabling conclusions on shape accuracy and region density.

[2]http://www.vision.caltech.edu/pmoreels/Datasets/TurntableObjects

The definition of region shape varies between detectors: Both HARAFF and HESAFF use the second moment matrix of the intensity gradient (Baumberg [1], Lindeberg and Garding [3]) to define an elliptic region description, EBR uses parallelograms derived from the edge geometry of the surrounding neighborhood, IBR regions are based on intensity profiles along rays emanating from an intensity ex-

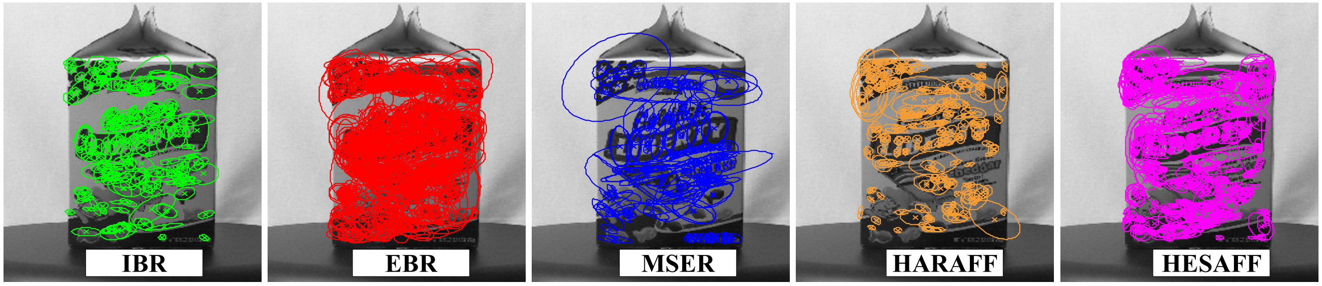

Figure 2. Correspondences between the *1st* and *2nd* frame of the *carton*-sequence are shown, overlayed on the *1st* frame. Region density corresponds to the ratio of positives and negatives in table 1.

tremum and MSER returns the boundaries of watershed-like regions. In order to enable meaningful comparisons between the different detectors, each region is replaced by an (approximated) ellipse as described in [7]. For elliptic regions, potential correspondences may be easily identified based on the *area overlap error*

$$e_o^{i,j}(m,n) = 1 - \frac{H_{ij} r_{i,m} \cap r_{j,n}}{H_{ij} r_{i,m} \cup r_{j,n}} , \qquad (1)$$

where $r_{i,m}$ and $r_{j,n}$ are corresponding regions in frames i and j. Further holds $m \in \{1...M\}$ and $n \in \{1...N\}$, where M and N denote the number of regions in either frame. H_{ij} represents the estimated ground truth homography between the frame pair. In the remainder of this work, both frame and region indices are omitted for the sake of brevity. Region congruency is reached for $e_o = 0$ whereas ellipses with no intersection yield $e_o = 1$. Two regions are associated to each other as candidate correspondences if $e_o \leq e_{o,max}$.

If several regions in frame i should claim the same candidate in frame j or vice versa, ambiguities have to be resolved based on the degree of e_o for each candidate correspondence (m, n). Figure 3 illustrates the principle: In the first step, the region pair with lowest overlap error is selected from an initial link set and all dependent correspondence candidates are removed. On the reduced set, this procedure is applied iteratively, until all candidates have either been labeled as final (*positives*) or rejected (*negatives*) correspondences.

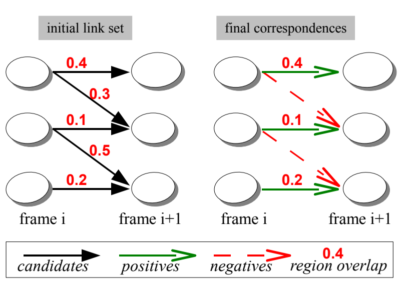

Figure 3. Iterative correspondence assignment.

In their comparative evaluation [7], Mikolajczyck *et al.* always chose the first image of each sequence as a reference frame in order to measure the influence of different geometric and photometric transformations on detector performance. Throughout this evaluation, correspondences are selected between adjacent frames instead. The commonality which all 6 sequences share is the mere observation that perspective transformations do occur, inflicting change in shape, scale, orientation and position upon all regions. In contrast to [7], neither type nor degree of the transformations are considered here.

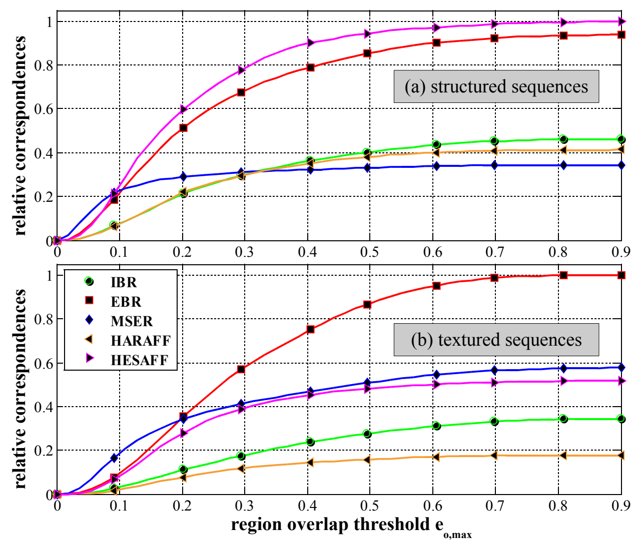

Figure 4. Number of correspondences (*positives*).

In order to obtain an impression on the susceptibility of each detector to variations of the maximum overlap threshold, figure 4 shows the relative number of positives as a function of $e_{o,max}$. In the literature [7][9], the setting $e_{o,max} = 0.5$ is normally used. In this work, the same threshold has been adopted, bounding the maximum localization error in subsequent sections. In figure 4 it can be seen, that all detectors converge to a maximum number of positives. For the structured sequences, EBR and HESAFF clearly detect the most correspondences, followed with distance by IBR, HARAFF and MSER. It is conspicuous that on the textured sequences, the detection rate of HESAFF is significantly lower compared to EBR and is even slightly outperformed by the MSER-detector. Also, differences between IBR and HARAFF are more pronounced.

From the curve progressions, a first assessment of shape accuracy is possible: With MSER for example, the number of positives with an overlap error below $e_o \leq 0.2$ has increased to more than 80% of all positives found for $e_o = e_{o,max}$ (for the structured sequences), indicating high region accuracy. In the case of IBR on the contrary, only slightly more than 50% of all positives have an overlap error of $e_o \leq 0.2$. Except for MSER, differences between the detectors are comparatively small. For the textured sequences, gradients are generally weaker and notably, there is almost no difference between IBR and EBR. In section 4, it will be shown that the actual localization errors generally coincide with these observations. Additionally, table 1 shows the absolute number of positives and negatives for each detector at $e_o = e_{o,max}$ in all frames of each sequence. From the number of negatives, conclusions on region density and thus matching complexity can be drawn: With EBR, the negatives exceed the positives by a factor of 5 on both structured and textured sequences, indicating high density. With MSER on the contrary, region density is lowest. Figure 2 gives a visual impression of region densities for all detectors.

		boat	wall	dvd	$\sum$	graffiti	carton	frame	$\sum, \varnothing$
EBR	$\sum pos.$	2036	5457	628	**8121**	2083	3428	419	**5930**
	$\sum neg.$	13062	26344	2173	**41579**	11062	19505	637	**31204**
	$\frac{neg.}{pos.}$	6.42	4.83	3.46	**5.12**	5.31	5.69	1.52	**5.26**
IBR	$\sum pos.$	969	1042	567	**2578**	1325	1142	325	**2792**
	$\sum neg.$	1248	725	596	**2569**	1573	1273	274	**3120**
	$\frac{neg.}{pos.}$	1.29	0.70	1.05	**0.99**	1.19	1.11	0.84	**1.12**
MSER	$\sum pos.$	1284	3127	348	**4759**	1056	1079	159	**2294**
	$\sum neg.$	473	1118	170	**1761**	705	547	55	**1307**
	$\frac{neg.}{pos.}$	0.37	0.36	0.49	**0.37**	0.67	0.51	0.35	**0.57**
HARAFF	$\sum pos.$	643	499	334	**1476**	915	1327	405	**2647**
	$\sum neg.$	1183	736	560	**2479**	1685	2759	682	**5126**
	$\frac{neg.}{pos.}$	1.84	1.47	1.68	**1.68**	1.84	2.08	1.68	**1.94**
HESAFF	$\sum pos.$	1644	1364	1510	**4518**	2044	3418	1109	**6571**
	$\sum neg.$	4796	2924	4340	**12060**	6267	9662	3047	**18976**
	$\frac{neg.}{pos.}$	2.92	2.14	2.87	**2.67**	3.07	2.83	2.75	**2.89**

Table 1. Number of correspondences (*positives*,top row), rejected candidates (*negatives*,middle row) and ratio of rejected and accepted candidates (bottom row) for the textured (left) and the structured sequences (right) at $e_o = e_{o,max} = 0.5$.

4. Localization Accuracy

The area overlap error in equation 1 is influenced by four different region properties: position of the region center, ratio of minor and major axis, direction of the major axis (on the interval $0 \dots \pi$) and region scale s (square root of major and minor axis). From e_o alone, conclusions on the individual error of a specific property may not be drawn. To this purpose, the position error has been additionally evaluated, as it is of great interest in many applications (*e.g.* homography estimation).

The *position error* is defined as the absolute euclidean distance

$$e_p^{i,j}(m,n) = \|H_{ij} p_{i,m} - p_{j,n}\| \qquad (2)$$

where $p_{i,m}$ and $p_{j,n}$ denote the positions of the corresponding region centers. As with e_o, both frame and region indices will be omitted for brevity.

4.1. Overlap-based correspondences

In this section, region correspondence assignment is based on the area overlap error as described in section 3. Therefore, the resulting evaluation of e_p and e_o represents only the theoretically achievable detector accuracy, as for real matching applications homographies H_{ij} between frames are usually not known and an overlap measure can thus not be computed.

Figure 5 shows the distribution of e_p over the n-percentile for $e_{o,max} = 0.5$: For the *structured sequences*, approximately 80 % of all MSER-correspondences show an error below $e_p \leq 2\ pel$ and only 5 % are worse than $e_p = 4\ pel$. Between HARAFF and HESAFF, there are almost no differences while IBR performs worst, closely followed by EBR. For the *textured sequences*, e_p is generally higher, but with almost identical relative ordering of the detectors (only EBR now is slightly worse than IBR). For MSER, 80 % of all correspondences show an error below $e_p \leq 3\,pel$. For higher percentiles, e_p is significantly higher than for the structured sequences. In figure 6, the region overlap error e_o over the n-percentile for $e_{o,max} = 0.5$ is shown. Again, MSER shows the highest accuracy, with an error below $e_o \leq 0.05$ for 80 % of all correspondences on the structured sequences. The other detectors may be classified into two groups: HARAFF/IBR show the worst performance, with 50 % of all correspondences above $e_o \geq 0.05$. The second group, HESAFF/EBR, is slightly better but still less accurate than MSER. In analogy to e_p, the overlap error for the textured sequences is generally higher than for the structured sequences. Notably, differences between the detectors are much less pronounced.

4.2. Accuracy as a function of scale

Considering a small circular region of scale $s = 10\ pel$ and a correspondence with $e_o = 0.5$ as illustrated in figure 7, the resulting position error is at $e_p = 5\ pel$. For a larger region with $s = 40\ pel$ and the same overlap error, the position error attains $e_p = 20\ pel$ however. Thus, a detector that produces mainly large regions is in principle more prone to larger position errors. For better comparability among different detectors, e_p could be normalized on region scale. However, this has not been done in the context of this evaluation, as information on the *true* error can be used more effectively as a decision criterion with regard to a specific target application. In figure 8, the distribution

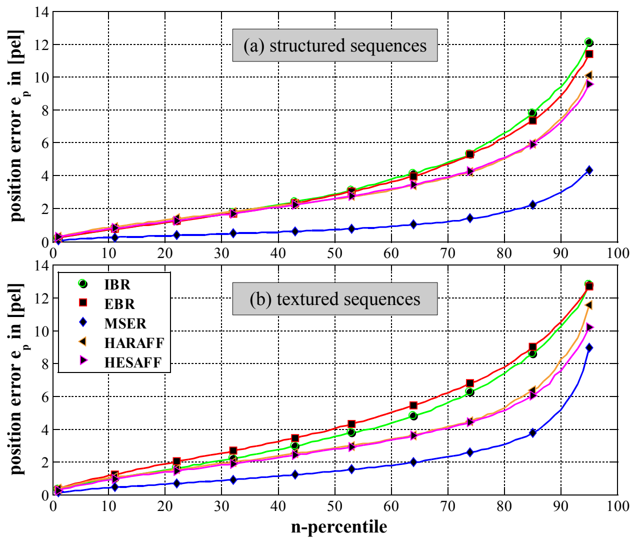

Figure 5. Distribution of the position error e_p.

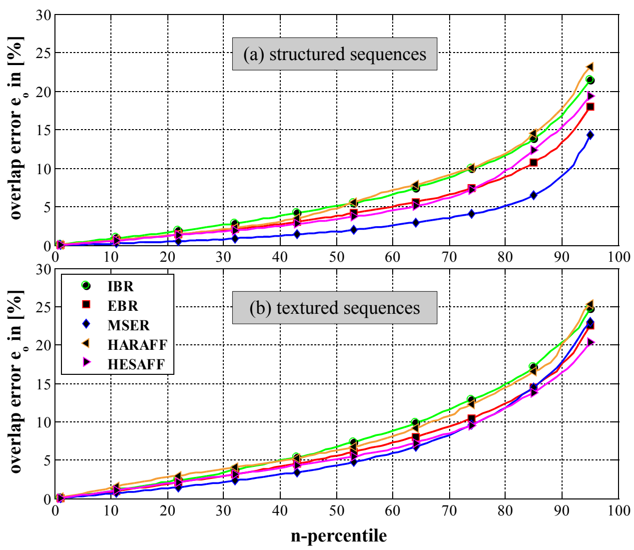

Figure 6. Distribution of the region overlap error e_o.

of region scales is shown for each detector. While IBR, MSER, HARAFF and HESAFF are very similar with more than 80 % of all regions below $s = 30\ pel$, EBR provides larger regions with more than 50 % above $s = 30\ pel$.

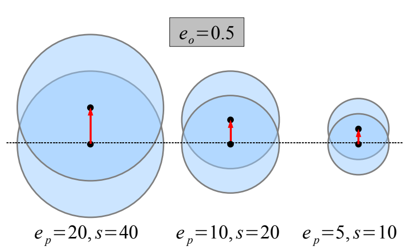

Figure 7. Influence of region scale s on position error e_p in [pel].

Notably, although EBR-regions are generally larger than IBR-regions, the position errors for both detectors are very similar. Also, e_p is significantly lower for MSER than for IBR, while both methods detect regions of comparable scales. Thus, the distribution of scales alone is not sufficient in order to explain detector differences. Therefore, the influence of region scale on both e_p and e_o has been additionally evaluated statistically in figure 9. Here, different percentiles of both errors are shown as a function of scale. For compactness, a separate evaluation of structured and textured sequences has been spared.

While e_p does increase linearly with scale for IBR, EBR, HARAFF and HESAFF, it remains largely constant for MSER. Strikingly, the latter exhibits a negative linear dependency of e_o instead: The median position error of MSER-regions decreases from $e_o \approx 0.1$ at $s = 10$ to only $e_o \approx 0.01$ at $s = 60\ pel$. In conjunction with figure 8, the distributions of e_p and e_o may be used as a pre-selection criterion in order to improve localization accuracy: By removing all regions with $s > 20\ pel$ for HARAFF and HESAFF for example, the median position error can be lowered to $e_p \approx 2\ pel$ while preserving approximately 60% of all correspondences.

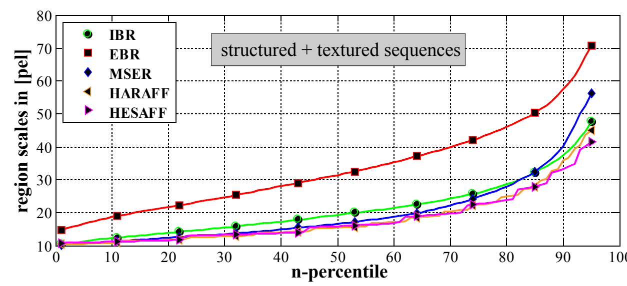

Figure 8. Distribution of region scales.

4.3. Descriptor-based correspondences

For the evaluation in section 4.1, regions have been assigned based on the area overlap error e_o. If instead region descriptors are used as selection criterion, a different set of correspondences results, generally with lower accuracy. Within this work, SIFT [4] has been used as a popular representative for the class of histogram-based region descriptors. The purpose of this section is two-fold: Firstly, it is investigated in how far both methods differ from each other with regard to set intersection and localization accuracy. Secondly, three measures for the mitigation of set differences and for the removal of error-prone correspondences are discussed. In order to reduce the number of figures within this publication, textured and structured image sequences are not evaluated separately. For the same reason, localization accuracy has been assessed in terms of e_o only.

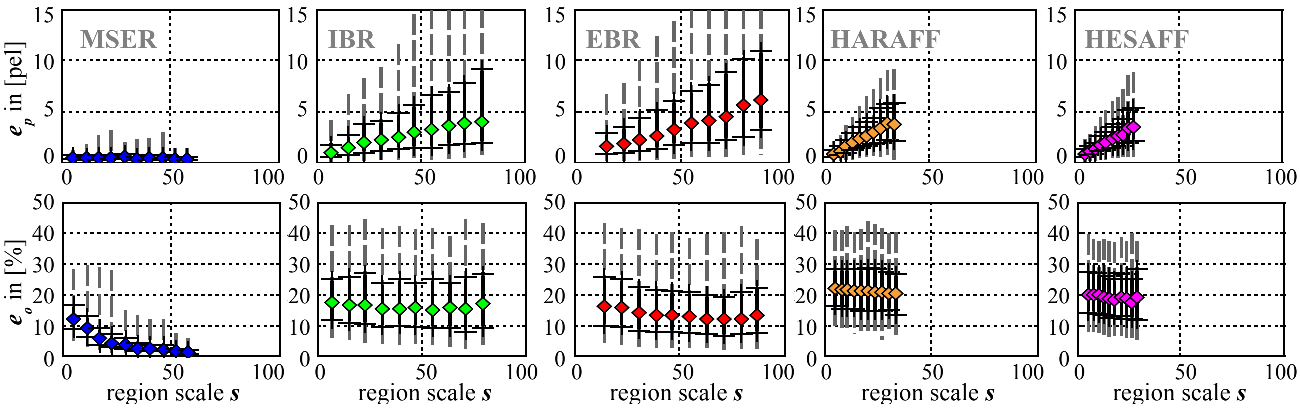

Figure 9. Dependency of region accuracy in terms of position error e_p (top) and overlap error e_o (top) on region scale s. The diagrams show the median error (diamond markers), the 25- and 75-percentiles (solid lines) and the 5- and 95-percentiles (dashed lines), estimated from all sequences (textured + structured).

In the following, the set of correspondences based on e_o is termed c_o and the set based on descriptor distance is referred to as c_d. Set differences are defined as $c_{d,d} = c_d \notin c_o$ and $c_{d,o} = c_o \notin c_d$. Figure 10 illustrates the terminology.

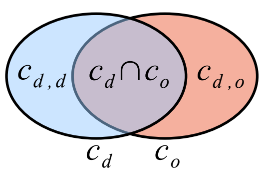

Figure 10. Set terminology.

Naturally, c_d and c_o may only differ in the subset of correspondences with multiple candidate regions: For regions with only a single assignment candidate, descriptor-based and overlap-based decisions are always identical. In figure 11, the distributions of e_o for $c_{d,d}$ (solid line) and $c_{d,o}$ (dashed line) are shown: With all detectors, e_o is lower for $c_{d,o}$. The number of correspondences with multiple candidates and the percentage of set differences $\frac{c_{d,d}}{c_d}$ are additionally given in table 2. Although for MSER the difference between $c_{d,d}$ and $c_{d,o}$ in figure 11 is most significant, it affects only 13 % of all correspondences in c_d. For EBR, differences in e_o between both sets are smaller, but the ratio $\frac{c_{d,d}}{c_d}$ is at 34 %, thus affecting a larger part of correspondences. The same relative ordering of the detectors as in table 2 can also be found in table 1 (section 3) in the ratio $\frac{positives}{negatives}$: In the case of EBR for example, the number of negatives exceeds the positives by a factor of 5, indicating high region density. As a consequence, there is an increased probability for picking different correspondences c_d, which explains the high percentage of set differences. For MSER, there are fewer negatives than positives and correspondingly, set differences are lowest among all detectors.

In order to reduce $c_{d,d}$ and thus to improve localization accuracy, several strategies for the removal of error-prone correspondences are discussed in the following. In the case of *nearest-neighbor matching*, two regions are assigned to each other if their descriptors d_m and d_n are best matches and the distance d between them is below a threshold:

	IBR	EBR	MSER	HARAFF	HESAFF	∅
$\sum$ *corresp.*	5370	14051	7053	4123	11089	**8337**
$\sum$ *cand.* ≥ 2	2324	10471	1650	2512	8410	**5073**
$\frac{c_{d,d}}{c_d}$	24%	34%	13%	27%	30%	**25.6%**

Table 2. Number of all correspondences (*top row*), correspondences with > 1 candidates (*middle row*) and percentage of set differences between overlap-based and descriptor-based assignment (*bottom row*).

$$d = \min\{d_{m,n} | d_{m,n} = \|d_m - d_n\| \leq d_{max}\} \qquad (3)$$

Figure 12a shows the distribution of descriptor distances d for the difference set $c_{d,d}$ and for the intersection $c_d \cap c_o$. It can be seen, that correspondences in $c_{d,d}$ have a significantly higher descriptor distance. According to the distribution, a threshold of $d = 0.30$ has to be selected in order to preserve 80 % of all correspondences in $c_d \cap c_o$ while $c_{d,d}$ is reduced by 50 %. For this setting, the expected region overlap error will be below $e_o \approx 0.2$ according to figure 12d with a spread of $\sigma_{e_o} \approx 0.05$. According to table 3, the new ratio $\frac{c_{d,d}}{c_d}$ after thresholding is at 14.2 % (compared to 25.6 % in table 2 when no threshold is used).

The second strategy is based on *descriptor uniqueness*, where thresholding is applied to the ratio

$$u_d = \frac{\|d_m - d_n\|}{\|d_m - d_o\|}, \qquad u_d \leq u_{max} \qquad (4)$$

with d_n and d_o as the best and second-best matches to d_m. The resulting distributions of u_d are shown in figure 12b. As with nearest-neighbor matching, a linear dependency exists between u_d and e_o. In order to preserve a similar number of correspondences, a threshold of $u_d = 0.75$ was selected. The expectable e_o and the spread σ_{e_o} according to

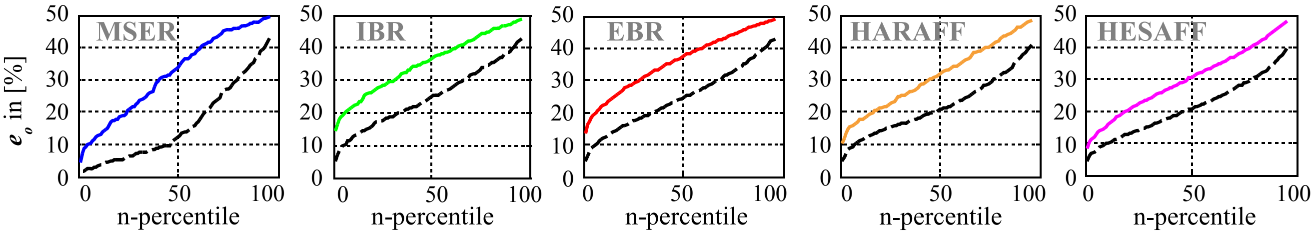

Figure 11. Distribution of region overlap errors e_o for descriptor-based correspondences ($c_{d,d}$,solid line) and overlap-based correspondences ($c_{d,o}$,dashed line).

figure 12e are almost identical to nearest-neighbor matching, while the ratio $\frac{c_{d,d}}{c_d}$ is reduced to 14.6%.

Thirdly, a new strategy - *shape uniqueness* - is introduced in this paper, which evaluates the geometrical overlap of neighboring regions. Using the definition of the area overlap error from equation 1, shape uniqueness u_s is defined as the minimum overlap between a region r_m and its neighbors within the same frame:

$$u_s = 1 - \min\{e_o^{i,i}(m,n) | e_o^{i,i}(m,n) \leq e_{o,max}\} , \quad (5)$$

where both $m, n \in \{1...M\}$ and M is the total number of regions. The resulting distributions of u_s can be seen in figure 12c. In order to obtain a similar number of correspondences as with nearest-neighbor matching and descriptor uniqueness, a threshold of $u_s = 0.75$ was selected. While e_o is almost identical to before, the spread σ_{e_o} as seen in figure 12f is significantly lower, especially for small values of u_s. This enables a more precise estimation of e_o for a given threshold on u_s. While the number of correspondences is roughly equivalent to the other strategies, the ratio $\frac{c_{d,d}}{c_d}$ is reduced to only 6.2%.

	IBR	EBR	MSER	HARAFF	HESAFF	∅
			$d = 0.30$			
$\sum$ *cand. corresp.*	11059	86834	10121	11728	42125	**32373**
$\sum$ *final corresp.*	1272	6047	1264	1649	5804	**3207**
$\frac{c_{d,d}}{c_d}$	10%	18%	5%	18%	20%	**14.2%**
			$u_d = 0.75$			
$\sum$ *cand. corresp.*	11059	86834	10121	11728	42125	**32373**
$\sum$ *cand.* ≥ 2	1565	6049	1308	1680	5557	**3232**
$\frac{c_{d,d}}{c_d}$	13%	18%	6%	18%	18%	**14.6%**
			$u_s = 0.75$			
$\sum$ *cand. corresp.*	8822	40013	8604	5676	15900	**15803**
$\sum$ *cand.* ≥ 2	1517	6216	1343	1633	5228	**3187**
$\frac{c_{d,d}}{c_d}$	5%	9%	2%	7%	8%	**6.2%**

Table 3. Number of candidate (*top rows*) and final correspondences with > 1 assignment candidates (*middle rows*), percentage of set differences between overlap-based and descriptor-based assignment (*bottom rows*) for all three selection methods.

The major advantage of this new measure is the possibility to apply it to each image individually prior to matching. Thereby, the number of correspondences and thus assignment complexity can be significantly reduced. As seen in table 3, this is especially advantageous in the case of high region densities, as with HESAFF or EBR. For the latter, 86834 candidate correspondences had to be evaluated with regard to descriptor similarity d or descriptor uniqueness u_d in order to obtain ≈ 6050 correspondences. By thresholding all regions with shape uniqueness $u_s \leq 0.75$ before the assignment step, the number of candidate correspondences could be reduced by almost 50% to 40013 while preserving a similar amount of final correspondences. For the three sets of final correspondences with more than 2 candidates, both e_p and e_o have been compared in figure 13. For all detectors, the set based on u_d shows the highest error both in e_p and in e_o. Except for e_o with EBR and MSER, shape uniqueness provides the highest localization accuracy.

5. Summary and Outlook

In this work, five popular region detectors have been compared with regard to localization accuracy in position and shape. Based on carefully estimated ground truth ho-

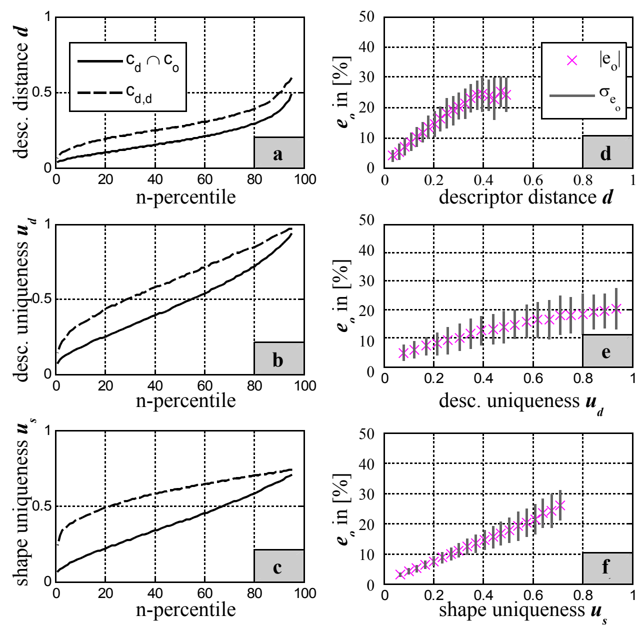

Figure 12. **Left:**Distribution of descriptor-based and shape-based measures for the removal of error-prone correspondences. **Right:** Dependencies between measures and region overlap error e_o.

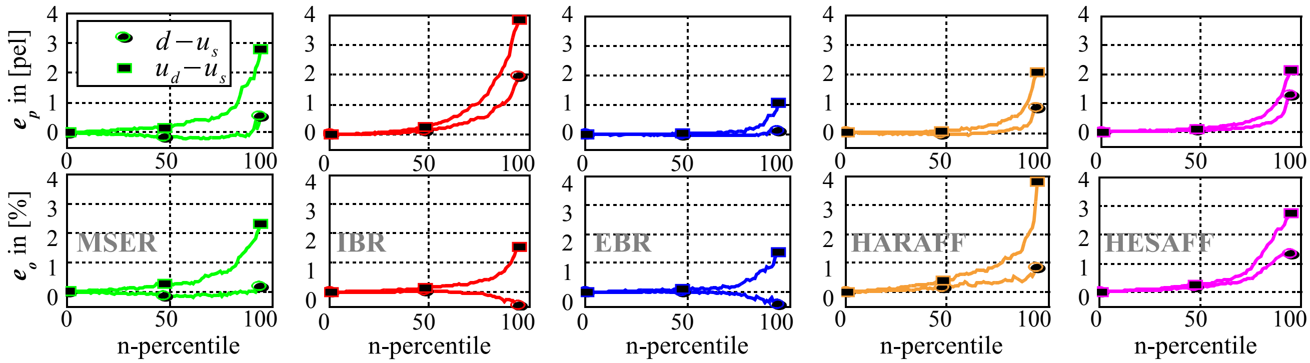

Figure 13. Localization accuracy in position (top) and scale (bottom). The curves show the differences between descriptor distance d, descriptor uniqueness u_d and the newly introduced shape uniqueness u_s.

mographies, correspondences between frames have been assigned based on the area overlap error. The image set consisted of 6 sequences categorized into two disjoint groups.

Concluding, the best detector with regard to both shape and position accuracy on all 6 sequences is MSER, followed with some distance by HESAFF. Considering only position accuracy, EBR and IBR performed worst. With regard to shape accuracy, EBR and HARAFF showed the highest errors. Except for HESAFF and HARAFF, localization accuracy in both position and shape was generally worse on the textured sequences. Further, it has been shown that for IBR, EBR, HARAFF and HESAFF, the position accuracy linearly depends on region scale. While the latter remains constant over scale for MSER, the overlap error has shown significantly higher for smaller regions. Based on these results, scale can be used as a pre-selection criterion for the removal of error-prone regions as a prerequisite to matching.

When descriptor distance is used for candidate assignment instead of area overlap, a different set of correspondences results with lower accuracy. For EBR, the differences between both methods are most significant, closely followed by HESAFF. It has been shown, that the percentage of set differences is related to region density, which is highest for EBR and HESAFF and lowest for MSER. In order to improve the accuracy of descriptor-based correspondences, three measures for the removal of error-prone regions have been discussed: While for both *descriptor distance* and *descriptor uniqueness* correspondences must be known, the newly introduced measure *shape uniqueness* is pre-computed on each image individually. Thus, the complexity of the matching task could be reduced by more than 50% on the investigated sequences. While the number of final correspondences is very similar for all three methods, set differences and localization error are lowest if shape uniqueness is used.

In future work, the affine salient region detector by Kadir *et al*. [2] will be included into the evaluation. Also, different region descriptors will be tested with regard to set differences such that a suitable descriptor may be chosen for each detector. Further, it will be investigated if shape uniqueness is a suitable measure for the automated adaptive configuration of detector parameters in real-time applications.

References

[1] A. Baumberg. Reliable feature matching across widely separated views. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 1:774–781, 2000.

[2] T. Kadir, A. Zisserman, and M. Brady. An affine invariant salient region detector. *European Conference on Computer Vision*, 1:228–241, 2004.

[3] T. Lindeberg and J. Garding. Reliable feature matching across widely separated views. *Image and Vision Computing*, 15:415–434, 1997.

[4] D. Lowe. Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision*, 60(2):91–110, 2004.

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

[6] K. Mikolajczyk and C. Schmid. Scale & affine invariant interest point detectors. *International Journal of Computer Vision*, 60(1):63–86, 2004.

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

[8] P. Moreels and P. Perona. Evaluation of features detectors and descriptors based on 3d objects. *Tenth IEEE International Conference on Computer Vision*, 1, 2005.

[9] C. Schmid, R. Mohr, and C. Bauckhage. Evaluation of interest point detectors. *International Journal of Computer Vision*, 37(2):151–172, 2000.

[10] T. Tuytelaars and L. Van Gool. Matching Widely Separated Views Based on Affine Invariant Regions. *International Journal of Computer Vision*, 59(1):61–85, 2004.