Morphological Distinguished Regions*

Allan Hanbury

Pattern Recognition and Image Processing Group (PRIP),
Institute of Computer Aided Automation, Vienna University of Technology,
Favoritenstraße 9/1832, A-1040 Vienna, Austria
hanbury@prip.tuwien.ac.at

Abstract. Distinguished regions can be detected with high repeatability in different images of the same scene. Two definitions of distinguished regions of an image in a mathematical morphology framework are proposed: one based on the use of reconstruction operators on a series of cross sections of a greyscale image, and the second based on extracting regions present in a large number of levels of a watershed segmentation hierarchy. The proposed distinguished regions are evaluated by measuring their repeatability in transformed images of the same scene.

Keywords: distinguished region, hierarchical segmentation, repeatability, watershed.

1 Introduction

Regions that can be detected with high repeatability in different images of the same scene or object have many uses in computer vision. They have been particularly useful for finding correspondences between images for wide-baseline stereo matching [1] and for locating features for object recognition [2]. These regions have been referred to as *invariant regions* [3], *covariant regions* [4] and *distinguished regions* [1]. We use the latter name in this paper.

We investigate distinguished regions that can be extracted within a mathematical morphology framework. The first part, in Section 2, is of more theoretical interest, as we show how a stricter version of the MSER detector [1] can be defined using reconstruction operators on a sequence of cross sections of a greyscale image, in a similar way to which the regional minima and maxima are defined.

We then consider extracting distinguished regions from segmentation hierarchies (Section 3). Hierarchies encoding the fusion of regions created during a watershed flooding of an image have been well studied [5]. These hierarchies have been used to create an image partition containing a specified number of regions by choosing a specific level of the hierarchy [6] or to assist in the manual creation of image partitions by allowing simple fusion and splitting of regions [7]. Nevertheless, the calculation of image features from the complete hierarchy instead of from a single level of the hierarchy has received little attention, showing that

* Partially supported by the EU Network of Excellence MUSCLE (FP6-507752) and the Austrian Science Foundation (FWF) under grant SESAME (P17189-N04).

L. Rueda, D. Mery, and J. Kittler (Eds.): CIARP 2007, LNCS 4756, pp. 202–211, 2007.

much information available in the hierarchy is ignored. We suggest one possible feature that can be extracted from such a hierarchy, namely the number of levels in which a region of the partition is present. Regions which are present in a large number of levels are considered good candidates for distinguished regions.

Measurements of the repeatability of the proposed distinguished regions across transformed images of the same scene are presented in Section 4.

2 Intensity-Based Distinguished Regions

Of the region detectors described in [4], two operate directly on the intensity values of a greyscale image, while the others make use of the detection of corner points in a scale space or operate on the entropy of the probability density function of intensities in an area. The two intensity-based detectors are the *Intensity Extrema-Based Region Detector* (IBR) [8] and the *Maximally Stable Extremal Region Detector* (MSER) [1].

The IBR begins by locating local intensity extrema in a series of smoothed images. The derivative of the intensity is then evaluated on rays emanating from each local extremum and the derivative extremum on each ray is found. Connecting the positions of these extrema produces an irregularly shaped region, which is replaced by an ellipse having the same shape moments.

An MSER is a connected component of an appropriately thresholded image [4]. Given a greyscale image f with integer greylevels, an *Extremal Region* $\mathcal{Q}$ is a connected component with the property that $\forall p \in \mathcal{Q}, q \in \delta^{(1)}(\mathcal{Q}) \setminus \mathcal{Q} : f(p) > f(q)$, or alternatively $f(p) < f(q)$. An extremal region is considered to be maximally stable (i.e. to be an MSER) if the following holds [1]: Let $\mathcal{Q}_1, \ldots, \mathcal{Q}_{i-1}, \mathcal{Q}_i, \ldots$ be a sequence of nested extremal regions, i.e. $\mathcal{Q}_i \subseteq \mathcal{Q}_{i+1}$. Extremal region $\mathcal{Q}_{i^*}$ is maximally stable iff $q(i) = |\mathcal{Q}_{i+\Delta} \setminus \mathcal{Q}_{i-\Delta}| / |\mathcal{Q}_i|$ has a local maximum at i^* ($|\cdot|$ denotes cardinality). Δ is a parameter of the method.

As pointed out in [1], the algorithm for locating MSERs is essentially identical to an efficient watershed algorithm [9]. For the watershed, the focus is on finding the positions where two catchment basins merge, whereas for the MSER detection algorithm, the focus is on the rate of change of the area of each catchment basin with increasing threshold. The connected components at which the rate of change of area with threshold is a minimum are chosen as MSERs.

A very strict version of this MSER detection algorithm, in which the connected components for which the rate of change of area with respect to threshold is zero, can be formulated as a morphological reconstruction based operator. We use the following notation (from [10]): $\mathrm{CS}_t(f)$ denotes the cross section of greyscale image f at level t (defined as the set of image pixels whose values are greater than or equal to t) and $R_g^\delta(f)$ denotes the morphological reconstruction by dilation of a mask image g from a marker image f. In a similar way to finding the regional maxima and

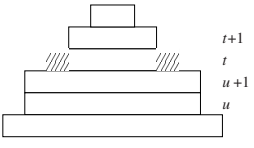


Fig. 1. Greyscale image cross sections

minima in [10], we can see that if a connected component is identical in $\mathrm{CS}_t(f)$ and $\mathrm{CS}_{t+1}(f)$, then it will not be reconstructed by $R^{\delta}_{\mathrm{CS}_t(f)}[\mathrm{CS}_t(f) \setminus \mathrm{CS}_{t+1}(f)]$ (see levels u and $u+1$ in Figure 1). If however the connected component in $\mathrm{CS}_t(f)$ is larger than the corresponding connected component in $\mathrm{CS}_{t+1}(f)$, then it will be reconstructed (see levels t and $t+1$ in Figure 1, where the hatches represent the set difference between the two levels). Hence the connected components at threshold level t that are identical in level $t+1$ can be found by

$$\mathrm{ST}_t(f) = \mathrm{CS}_t(f) \setminus R^{\delta}_{\mathrm{CS}_t(f)}[\mathrm{CS}_t(f) \setminus \mathrm{CS}_{t+1}(f)] \tag{1}$$

The binary image containing all the distinguished regions is calculated as the union of $\mathrm{ST}_t(f)$ taken over all greylevels present in the image, or

$$\mathrm{MDR}(f) = \bigcup \mathrm{ST}_t(f) \tag{2}$$

For images which do not contain pixels which assume all possible greylevel values, the values of t should be restricted to the greylevels for which the value of the greylevel histogram $H_t(f)$ is non-zero. This restriction is necessary as $\mathrm{CS}_t(f)$ will be identical to $\mathrm{CS}_{t+1}(f)$ if no pixels with a greylevel of t exist in the image, resulting in all connected components of $\mathrm{CS}_t(f)$ becoming distinguished regions.

Unfortunately it is not possible to find the stable connected components by a single greyscale reconstruction, as is done for the morphological regional maxima and minima, because the structure of a function as a stack of binary cross sections is not kept by the marker image in Equation 1. It may be possible to incorporate the idea of a minimum rate of area change from the MSER calculation through the use of attribute openings, but this remains to be investigated.

In practice, as the result of this algorithm includes very many small connected components, we apply an area opening of size λ to the result:

$$\mathrm{AMDR}(f) = \gamma_\lambda(\mathrm{MDR}(f)) \tag{3}$$

Figure 2(a) shows the regions obtained for an image with $\lambda = 25$ pixels.

3 Hierarchical Watershed-Based Distinguished Regions

We propose distinguished regions calculated from a watershed hierarchy based on volume extinction values [11,5]. The hierarchy is built as follows [5]. During the flooding process on an image, when a lake in a catchment basin is about to overflow, the dissimilarity between this catchment basin and its neighbour into which it would overflow is defined as the measurement of the volume of the full lake. This can easily be represented on a region adjacency graph (RAG), where each node represents a catchment basin, and each edge encodes the dissimilarity between two neighbouring catchment basins. This type of flooding is most often used to obtain a segmentation containing a specific number of regions: in order to obtain a partition with N regions, the $N-1$ edges with the highest dissimilarity values in the minimum spanning tree (MST) of the RAG are cut. Alternatively,

(a) Rec AO25 (b) Ellipses fitted to Rec AO25

(c) HWS Top50 (d) Ellipses fitted to HWS Top50

Fig. 2. (a) Outlines (in white) of the reconstruction-based morphological distinguished regions after an area opening of size 25 pixels. (c) Outlines of the 50 regions having the highest survival values. (b), (d) Ellipses fitted to the regions.

the hierarchy can be visualised as a stack of nested partitions, where the partition at level i contains i regions. This implies that in moving from level i to level $i+1$, one of the regions in level i is split into two regions in level $i+1$.

An obvious candidate for a distinguished region is a region which remains constant over a large number of levels of the hierarchy of partitions. We define the *survival value* of a region which "appears" at some level i (due to a split of a region at level $i-1$), and "disappears" by splitting into two at level $i+n$ to be n. If these survival values are encoded into the regions in each level of the hierarchy, then it is straightforward to find the longest surviving regions by searching for the largest survival values. For efficient searching, the regions in level i which are identical to regions in level $i-1$ are assigned survival values of 0. This avoids having a series of identical regions in subsequent levels with stepwise decreasing survival values, which would lead to multiple detection of a single region. On each level of the hierarchy (except for level 1), there will therefore be two regions with non-zero survival values. These two survival values on the same level are never identical, as both new regions cannot split at the same level. The 50 regions with the highest survival values in a hierarchy of 750 levels are shown in Figure 2(c).

4 Evaluation

We first summarise the framework for evaluating distinguished region repeatability. A description of the experiments and a discussion of the results follows.

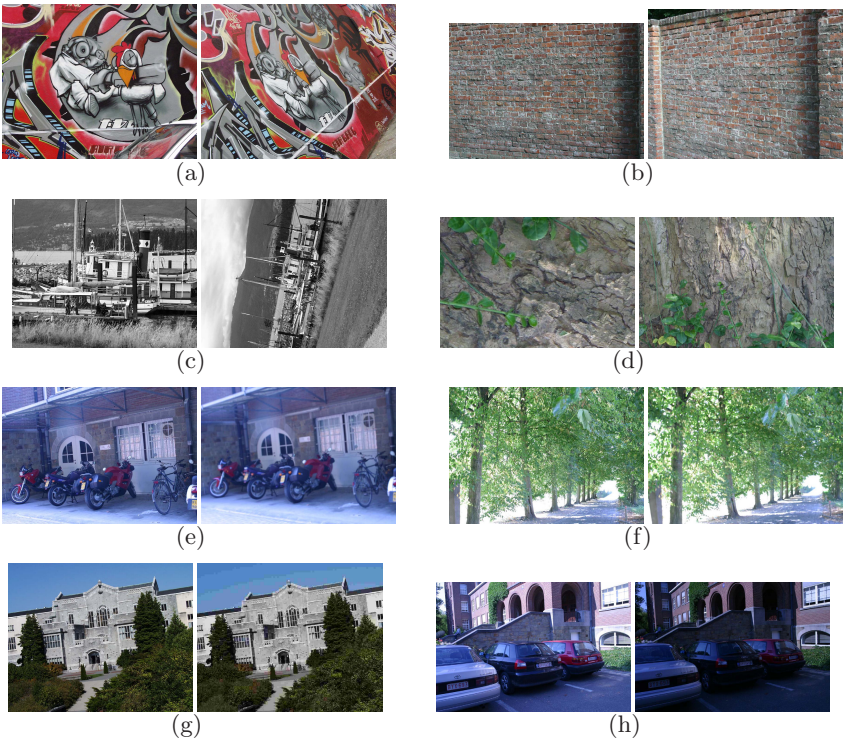


Fig. 3. Part of the evaluation dataset. (a), (b) Viewpoint change, (c), (d) Zoom+rotation, (e), (f) Image blur, (g) JPEG compression, (h) Light change. The leftmost image of each set is the reference image (from [4]).

4.1 Evaluation Framework

The *repeatability* measures the extent to which regions detected in transformed images of the same scene overlap. We use the evaluation framework presented in [4][1]. The framework consists of eight images, where each image is subjected to five transformations, resulting in sets of six images. Examples from the image sets are shown in Figure 3. The homographies between the reference images and the other images for each set have been computed, allowing the overlap between distinguished regions in the reference and another image to be evaluated.

In [4], only elliptical distinguished regions are considered, as these are intrinsically produced by four of the six algorithms tested. For the other algorithms, ellipses approximating the regions are chosen. To be compatible with the framework, we also fit ellipses to the edges of the regions produced by the proposed methods, using the ellipse fitting algorithm in [12]. Examples of the ellipses fitted to the detected regions are shown in Figures 2(b) and (d).

The repeatability is measured between the reference image and another image from the set. The distinguished regions are detected in both images and those

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

from the second image are projected onto the reference image by using the known homography. Two regions are said to form a region-to-region correspondence if the *overlap error* is sufficiently small — in this paper we use 0.4 as was done for the experiments in [4]. The overlap error is defined as [4]

$$1 - \frac{R_{\mu_a} \cap R_{(H^T \mu_b H)}}{R_{\mu_a} \cup R_{(H^T \mu_b H)}} \tag{4}$$

where R_u is the region enclosed by the ellipse defined by $x^T \mu x = 1$ and H is the homography relating the images. The *repeatability score* for a pair of images is the ratio between the number of region-to-region correspondences and the smaller number of regions in the pair of images. Only regions located in the part of the scene present in both images are counted. In addition, the regions are transformed to have a normalised size before calculating the overlap, to avoid the problems with regions of different sizes discussed in [4].

4.2 Experiments

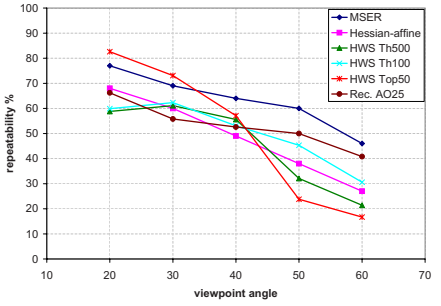
The results of the repeatability tests for the eight groups of six images in the dataset are shown in Figures 4 and 5, where the left column shows the repeatability percentage and the right column the number of correspondences. Curves corresponding to six methods are shown in each graph. The curves labelled *MSER* and *Hessian-affine* correspond to the two best performing methods of the six tested in the evaluation of affine covariant region detectors in [4].

For the distinguished regions based on the hierarchical watershed, the number of levels of each hierarchy is set to 750 (limited by the amount of memory on the computer used to perform the experiments). Each image is pre-processed by applying a leveling [13] of size 3 to each channel separately. The hierarchy is built on a gradient image obtained by applying the *saturation weighing-based colour gradient* in the L1 colour space [14]. We evaluate three methods of choosing the distinguished regions. The first two are based on setting a threshold on the minimum number of levels that a region must survive in order to be chosen as a distinguished region. We compared the use of a high threshold of 500 levels (*HWS Th500*) and a low threshold of 100 levels (*HWS Th100*). This evaluates if regions that survive over many levels in the original image also do so in the transformed images, or if choosing a lower threshold, thereby including more regions, leads to higher repeatability. The third method once again tests the repeatability of the regions which survive over many levels by choosing the 50 regions with the highest survival values (*HWS Top50*).

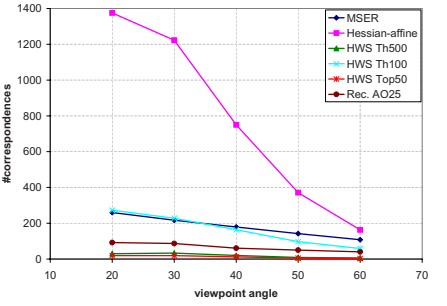
For the reconstruction based distinguished regions (Equation 3), we plot the results obtained using an area opening of size 25 pixels (*Rec. AO25*). Images were converted to greyscale before applying this method.

4.3 Discussion

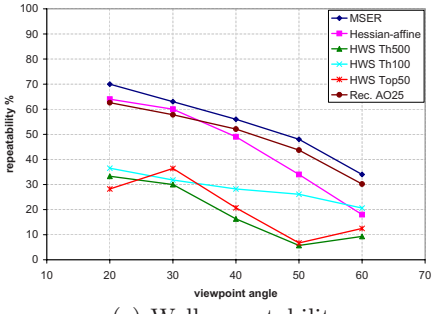
As has already been pointed out in [4], different algorithms perform better for different transformations, as can be seen by the repeatability results for the



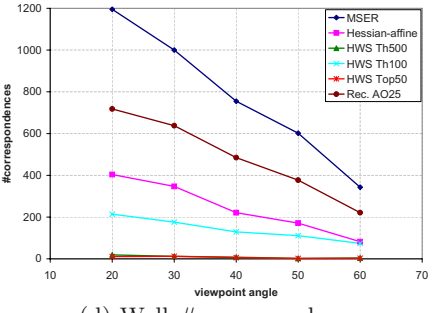
(a) Graf repeatability



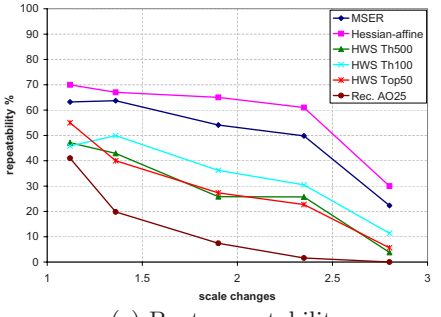
(b) Graf #correspondences



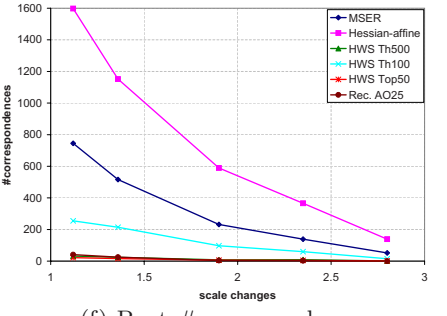
(c) Wall repeatability



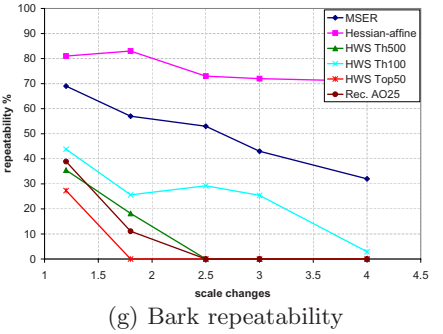
(d) Wall #correspondences



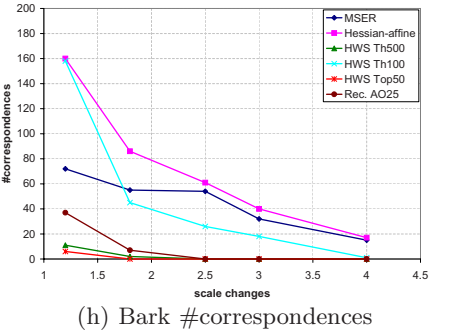
(e) Boat repeatability



(f) Boat #correspondences

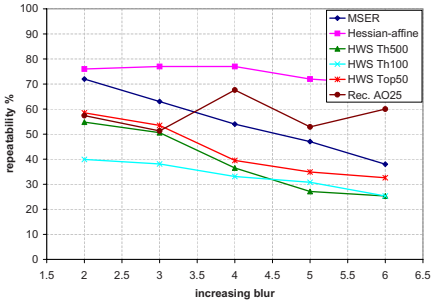


(g) Bark repeatability

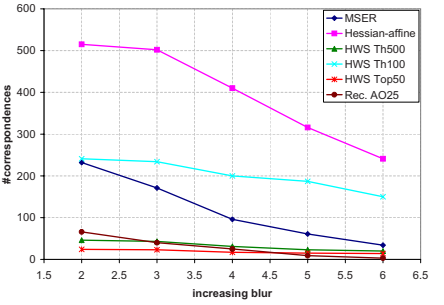


(h) Bark #correspondences

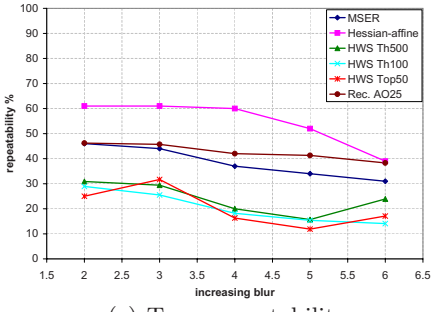
Fig. 4. The repeatability (left column) and number of correspondences (right column) for (a)–(d) viewpoint change and (e)–(h) scale change



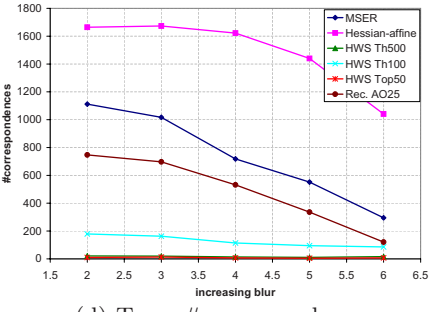
(a) Bikes repeatability



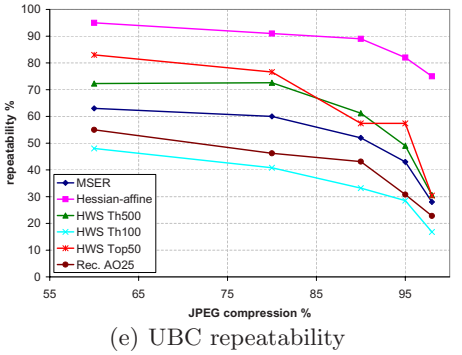
(b) Bikes #correspondences



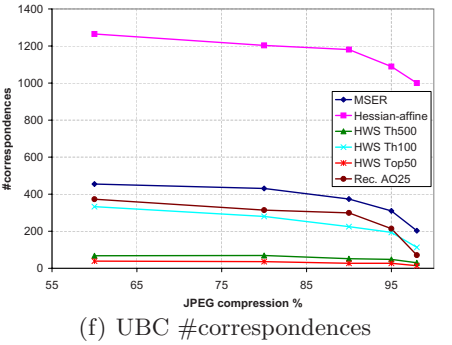
(c) Trees repeatability



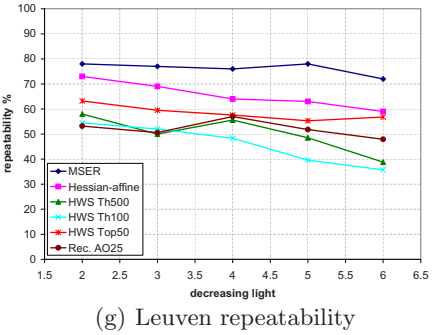
(d) Trees #correspondences



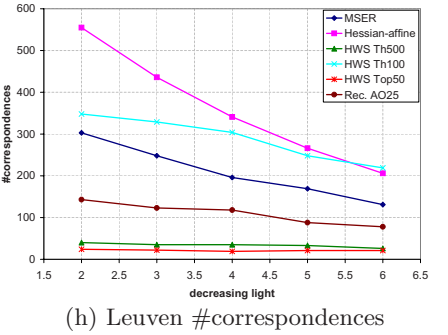
(e) UBC repeatability



(f) UBC #correspondences



(g) Leuven repeatability



(h) Leuven #correspondences

Fig. 5. The repeatability and number of correspondences for (a)–(d) blur, (e)–(f) JPEG compression and (g)–(h) illumination change

MSER and Hessian-affine detectors. The problem of comparing detectors producing different densities of regions is discussed in [4]. They point out that for detectors that produce few regions, the thresholds can be set so that the performance is often better than average. For detectors that produce many regions, the image may be so cluttered with regions that some get matched by accident. For the distinguished regions from the hierarchical watershed, we examine this effect by using the three ways listed above of choosing the number of distinguished regions. The Hessian-affine detector produces the largest number of correspondences for each image sequence except Wall, indicating that the density of the distinguished regions is the highest.

For the viewpoint changes (Figure 4(a)–(d)), the MSER detector has the highest repeatability. For the Graf image, it is interesting that the 50 regions with the highest survival value (*HWS Top50*) are extremely stable for small viewpoint changes and perform the worst for large viewpoint changes. The reconstruction based detector (*Rec. AO25*) performs similarly to the Hessian-affine detector for these images, outperforming it for the three largest viewpoint changes.

The morphological detectors prove to be bad at handling scale changes (Figure 4(e)–(h)). In particular for the Bark image, no matching distinguished regions were found by three of the methods as the zoom-out becomes markedly larger, leading to zero repeatability. Because we are dealing with scale changes, the corresponding regions will occur at different levels of the watershed hierarchy for different images. The fact that the (*HWS Th100*) performs the best among the tested algorithms demonstrates this, as the (*HWS Th500*) and (*HWS Top50*) only choose regions from the lower part of the hierarchy. Increasing the maximum number of levels in the hierarchy should improve the repeatability for scale change. For the blur images (Figure 5(a)–(d)), the reconstruction based approach produces better repeatability results than the MSER for 8 of 10 transformed images, but for a lower number of correspondences.

It is interesting that only for the JPEG compression images (Figure 5(e)–(f)), the two hierarchical watershed based methods in which few distinguished regions are selected perform better than the MSERs. Finally, for the illumination changes (Figure 5(g)–(h)), all morphological methods have a repeatability below that of the MSER and Hessian-affine methods.

In general, the *HWS Th500* and *HWS Top50* curves are very similar, indicating that there is little difference due to these methods of choosing the distinguished regions with the highest survival values. Based on the results, one cannot determine whether it is better to choose a lower survival value threshold (*HWS Th100*) leading to many distinguished regions or a higher threshold, as one of the thresholds does not lead to consistently better results over all transformations.

5 Conclusion

Two methods for calculating distinguished regions in a mathematical morphology framework are proposed. The first is based on reconstructions on a series

of cross sections of a greyscale image and the second extracts regions from a hierarchy calculated using a watershed based on volume extinction values.

Experiments measuring the repeatability of the extracted regions for different types of image transformations are presented. The repeatability falls into the range of the repeatability of the six algorithms tested in [4], without surpassing the best algorithms. One of the drawbacks of the evaluation framework used is that the difference in the number of regions (region density) extracted by each algorithm is not taken into account, which could affect the repeatability results.

The limit of 750 levels of the watershed hierarchy limits the performance, especially for scale changes. The advantage of the hierarchical approach is the extensive information available in the hierarchy for defining distinguished regions. We have so far only looked at one possible feature: the survival value. Use of information on the region shape, region inclusion information, neighbourhood information, etc., is an interesting area for future research.

References

1. Matas, J., Chum, O., Urban, M., Pajdla, T.: Robust wide-baseline stereo from maximally stable extremal regions. Image and Vision Computing 22, 761–767 (2004)
2. Csurka, G., Dance, C.R., Fan, L., Willamowski, J., Bray, C.: Visual categorization with bags of keypoints. In: Workshop on Statistical Learning in Computer Vision (at ECCV) (2004)
3. Dorko, G., Schmid, C.: Selection of scale invariant neighborhoods for object class recognition. In: Proc. of the Int. Conf. on Computer Vision, pp. 634–640 (2003)
4. Mikolajczyk, K., Tuytelaars, T., Schmid, C., Zisserman, A., Matas, J., Schaffalitzky, F., Kadir, T., Van Gool, L.: A comparison of affine region detectors. International Journal of Computer Vision 65(1/2), 43–72 (2005)
5. Meyer, F.: An overview of morphological segmentation. International Journal of Pattern Recognition and Articial Intelligence 15(7), 1089–1118 (2001)
6. Hanbury, A., Marcotegui, B.: Waterfall segmentation of complex scenes. In: Narayanan, P.J., Nayar, S.K., Shum, H.-Y. (eds.) ACCV 2006. LNCS, vol. 3851, pp. 888–897. Springer, Heidelberg (2006)
7. Zanoguera, F., Marcotegui, B., Meyer, F.: A toolbox for interactive segmentation based on nested partitions. In: Proc. of the Int. Conf. on Image Processing (1999)
8. Tuytelaars, T., Van Gool, L.: Matching widely separated views based on affine invariant regions. International Journal on Computer Vision 59(1), 61–85 (2004)
9. Vincent, L., Soille, P.: Watersheds in digital spaces: an efficient algorithm based on immersion simulations. IEEE Trans. Pattern Analysis and Machine Intelligence 13(6), 583–598 (1991)
10. Soille, P.: Morphological Image Analysis, 2nd edn. Springer, Heidelberg (2002)
11. Vachier, C., Meyer, F.: Extinction values: A new measurement of persistence. In: Proc. of the IEEE Workshop on Non Linear Signal/Image Processing, pp. 254–257 (1995)
12. Fitzgibbon, A.W., Pilu, M., Fisher, R.B.: Direct least-squares fitting of ellipses. IEEE Trans. Pattern Analysis and Machine Intelligence 21(5), 476–480 (1999)
13. Meyer, F.: Levelings, image simplification filters for segmentation. Journal of Mathematical Imaging and Vision 20, 59–72 (2004)
14. Angulo, J., Serra, J.: Color segmentation by ordered mergings. In: Proc. of the Int. Conf. on Image Processing. vol. II, pp. 125–128 (2003)