Research Article

Entropy-Based Maximally Stable Extremal Regions for Robust Feature Detection

Huiwen Cai,[1] Xiaoyan Wang,[2] Ming Xia,[2] and Yangsheng Wang[1]

[1] *Digital Interactive Media Laboratory, Institute of Automation, Chinese Academy of Sciences, Beijing 100190, China*
[2] *College of Computer Science and Technology, Zhejiang University of Technology, Hangzhou 310023, China*

Correspondence should be addressed to Huiwen Cai, huiwen.cai@ia.ac.cn

Maximally stable extremal regions (MSER) is a state-of-the-art method in local feature detection. However, this method is sensitive to blurring because, in blurred images, the intensity values in region boundary will vary more slowly, and this will undermine the stability criterion that the MSER relies on. In this paper, we propose a method to improve MSER, making it more robust to image blurring. To find back the regions missed by MSER in the blurred image, we utilize the fact that the entropy of probability distribution function of intensity values increases rapidly when the local region expands across the boundary, while the entropy in the central part remains small. We use the entropy averaged by the regional area as a measure to reestimate regions missed by MSER. Experiments show that, when dealing with blurred images, the proposed method has better performance than the original MSER, with little extra computational effort.

1. Introduction

Maximally stable extremal regions (MSER) [1] is a popular local invariant feature detection method because it has many highly desirable properties, such as invariance to monotonic intensity transformation, invariance to adjacency preserving (continuous) transformation, and low computational complexity [2–4]. It is shown in a detailed study [5] that MSER has the best repeatability and accuracy, except when dealing with blurred image. For blurred images, some local regions may no longer be detected by MSER, because MSER only detects regions whose shape is globally stable with respect to intensity perturbation. That is to say, the intensity difference should be large between pixels lying on the boundary and the outside. In the blurred image, sharp intensity contrast may be transformed into gradual variation. As
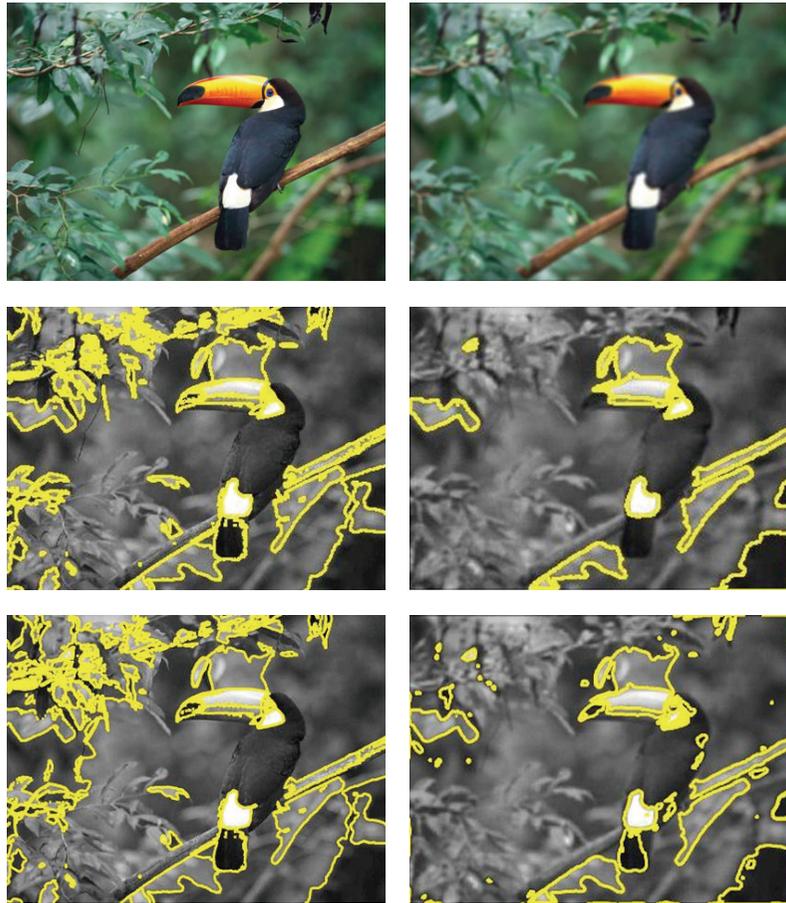
Figure 1: Boundaries of the detected blobs in images. The two on the top are the images for processing. The two in the middle are results of the MSER method. The two in the bottom are results of our method. The left ones are original images, while the right ones are the blurred versions.

demonstrated on the top of Figure 1, with MSER, some regions are lost in the blurred image due to the lack of sharp boundary, like the mouth and tail of the bird.

A blurred image can be seen as an image in a larger scale [6, 7]. Local feature detectors, which are good at dealing with blurring, like the Hessian affine detector and edge-based regions method, all rely on automatic scale selection methods. They construct a measure of local structure within the region and select the local maxima over the scales as the characteristic scale of the region. Even when the image is blurred, proper integration scale and differential scale can still be selected, the same structure can be found, not depending on intensity difference to find the boundary. A multiscale approach is used in [8] to improve MSER's ability to detect features in blurred images, which is a close concept to automatic scale selection. A shortcoming in both methods is that they need to consider many scales in each image position, which greatly increases the computational complexity [9–11]. Moreover, when measuring the structure within the image, a further scale selection may be needed to compute derivatives, such as the Laplacian operator in the Hessian affine detector. Another way to make MSER workable on blurred image is to relax its globally stable criteria, as done in

[12]. Stability is required only locally, for the primitives constituting the three-point frames. Higher number of features and better covered image are achieved, at the expense of more computational effort.

To enhance MSER's ability in the situation of blurring, we can still borrow the idea of using local image structure as a measure to select the proper region scale. In a blurred image, considering MSER's discarded extremal regions which are retained in original image before blurring, we can find a common phenomenon—intensities on the boundary vary more slowly compared to the original image, but the intensities in the central part are kept almost unchanged. In other words, entropy is large on the boundary and small inside. These extremal regions are lost because they are considered not stable with respect to intensity perturbation. To find these regions, we just need to find the extremal regions with minimal entropy. An advantage using entropy is that it does not rely on intensity difference, which may be changed in a blurred image. Another advantage is that entropy measures the region as a whole, instead of just one or a few pixels, which may be easily affected by blurring.

We just use the nested extremal regions extracted in the process of MSER for comparison of entropy, so that there is no additional effort for constructing series of regions of different scales, as what is done in automatic scale selection. Utilizing the nested extremal regions also facilitates the computation of entropy, which will be detailed in Section 2.

In the proposed method, we aim at finding the largest region with the minimal entropy. We use average entropy as our measure and find the region with the lowest average entropy within a local range in the nested extremal regions.

There is another popular local feature detector, "salient region" [13], using the entropy as a criterion. Different from our method, salient region looks for the smallest regions with maximal entropy, because it believes that these regions may contain more information. However, considering blurring, largest region with minimal entropy may be more proper. This is because, after blurring, entropy of region is inclined to increase. Therefore, region scale with maximal entropy will increase but scale with minimal entropy will decrease [14–17]. This could make the salient regions confusable with the background, but regions with minimal entropy are unaffected.

2. Entropy-Based MSER

The proposed method is an extension of the original MSER method, adding entropy as another criterion to find back extremal regions missed by MSER due to blurring. Here is the whole algorithm. Steps (1)–(4) describe the original MSER method, and step (5) is our additional step.

(1) Sort all the pixels by intensity.

(2) Place pixel one by one (in intensity order) in the image, and update the connected component structure, which forms the nested extremal regions.

(3) Compute the area variation for each extremal region:

$$\text{var}_i = \frac{\text{Area}_{i+\Delta} - \text{Area}_i}{\text{Area}_i}. \tag{2.1}$$

Figure 2: Comparison on blurred images. Here boundaries of the detected blobs are shown. The middle six are results of the MSER method. The right six are results of our method.

Here i represents the extremal region with maximal intensity i, and $i + \Delta$ refers to the extremal region expanded from the ith one, with maximal intensity $i + \Delta$. var_i is the relative difference of area when maximal intensity rises from i to $i + \Delta$.

(4) Traverse the nested extremal regions. Find the maximally stable extremal one, which has "var" smaller than its immediate parent and ancestor in the nest.

(5) For the extremal regions not considered as maximally stable, compute their entropy averaged by area. If an extremal region's average entropy is a local minimal in the neighborhood of the nest, we will select it as a complement to the MSERs.

It should be noted that the entropy of an extremal region can be easily calculated, because extremal regions are formed by sorted intensity; thus only subset of intensity is under consideration for each region. Here we show how to compute the average entropy in an incremental way.

Considering R_l as an extremal region with maximal intensity l and its direct parent in the nest R_{l+k}. R_l's intensity histogram can be represented as a_i $(i = 1, 2, \ldots, l)$. Because R_{l+k} contains R_l, besides a_i, there is additional intensity histogram b_i $(i = 1, 2, \ldots, l, l+1, \ldots, l+k)$. Entropy and area can be calculated as follows.

(1) Area of R_l: $S_l = \sum_{i=1}^{l} a_i$.

(2) Area of R_{l+k}: $S_{l+k} = \sum_{i=1}^{l} a_i + \sum_{i=1}^{l+k} b_i = S_l + \sum_{i=1}^{l+k} b_i$.

(3) Entropy of R_l: $H_l = -\sum_{i=1}^{l}(a_i/S_l)\log(a_i/S_l)$.

(4) Entropy of R_{l+k}: $H_{l+k} = -\sum_{i=1}^{l}((a_i + b_i)/S_{l+k})\log((a_i + b_i)/S_{l+k}) - \sum_{i=l+1}^{l+k}(b_i/S_{l+k})\log(b_i/S_{l+k})$.

We can see that, with intensity growing up, maximal intensity, area, and entropy all increase. Considering a simple case which often occurs, $b_i = 0$ $(i = 1, 2, \ldots, l)$, which means that the parent region R_{l+k} does not contain other subregions except R_l. Then there is a simple relationship between the average entropy:

$$\frac{H_{l+k}}{S_{l+k}} - \frac{\log S_{l+k}}{S_{l+k}} = \left(\frac{H_l}{S_l} - \frac{\log S_l}{S_l}\right) \times \left(\frac{S_l}{S_{l+k}}\right)^2 - \left(\frac{\sum_{i=l+1}^{l+k} b_i \log b_i}{S_{l+k}^2}\right). \tag{2.2}$$

It is easy to see that we can infer H_{l+k}/S_{l+k} from H_l/S_l in a simple way.

3. Experiments

In order to demonstrate the repeatability and time performance of our method when dealing with blurring, we compare it with the original MSER method, using open-source code provided at [18]. We use the 6 blurred images (bike set) provided by the popular Oxford dataset [19] for test. This blurred sequence is acquired by varying the camera focus [20, 21].

3.1. Repeatability Test

Here we show the results on the images for better comparison, as in Figure 2. We can see from Figure 2 that the main detection difficulties lie on the part where the intensity of the foreground is similar to the background, like the parts of the bike, and blurring makes

Table 1: Time comparison between the original MSER and our method.

Detection method	Computation time (seconds)
Original MSER	0.6476
Entropy-based MSER	1.3799

things worse. Results show that our entropy-based MSER could successfully detect this kind of regions, which will consequentially improve detection repeatability. Furthermore, the number of detected features is greatly increased, which is desirable in the matching and recognition stages.

3.2. Time Test

It is also useful to compare the time consumption between the original method and our method. Table 1 shows the time consumption on the above bike set on a 2.4 GHz Pentium 2 CPU. We can see that no significant extra time is needed in our method.

4. Conclusion

In this paper we have proposed an efficient extension to the maximally stable extremal regions (MSER) which makes it more robust to blurring. We utilize the fact that, in a blurred image, the entropy of probability distribution function of intensity values increases rapidly when the local region expands across the boundary, while the entropy in the central part keeps small, and use entropy averaged by region area as a measure to reestimate region missed by MSER. Experiments show that when dealing with blurred images, our method has better performance than the original MSER, with little extra computational effort.
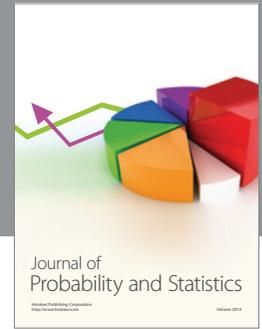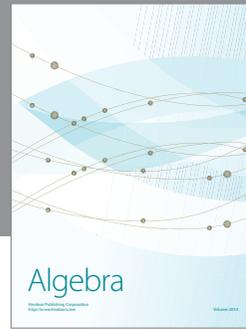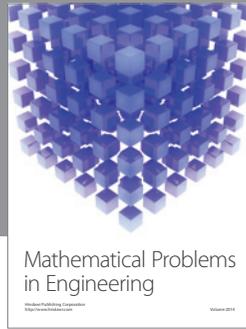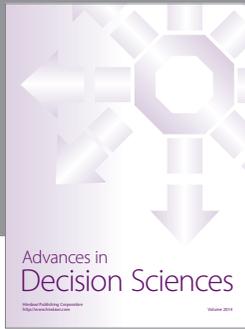
Acknowledgments

References

[1] J. Matas, O. Chum, M. Urban, and T. Pajdla, "Robust wide-baseline stereo from maximally stable extremal regions," *Image and Vision Computing*, vol. 22, no. 10, pp. 761–767, 2004.

[2] C. Cattani, "Shannon wavelets for the solution of integrodifferential equations," *Mathematical Problems in Engineering*, vol. 2010, Article ID 408418, 22 pages, 2010.

[3] S. Y. Chen, G. J. Luo, X. Li, S. M. Ji, and B. W. Zhang, "The specular exponent as a criterion for appearance quality assessment of pearl-like objects by artificial vision," *IEEE Transactions on Industrial Electronics*, vol. 59, no. 8, pp. 3264–3272, 2012.

[4] X. Zhang, Y. Zhang, J. Zhang et al., "Unsupervised clustering for logo images using singular values region covariance matrices on Lie groups," *Optical Engineering*, vol. 51, no. 4, Article ID 047005, 8 pages, 2012.

[5] K. Mikolajczyk, T. Tuytelaars, C. Schmid et al., "A comparison of affine region detectors," *International Journal of Computer Vision*, vol. 65, no. 1-2, pp. 43–72, 2005.

[6] J. Zhang, S. Y. Chen, S. Liu, and Q. Guan, "Normalized weighted shape context and its application in feature-based matching," *Optical Engineering*, vol. 47, no. 9, Article ID 097201, 2008.

[7] S. Y. Chen, H. Tong, Z. Wang, S. Liu, M. Li, and B. Zhang, "Improved generalized belief propagation for vision processing," *Mathematical Problems in Engineering*, vol. 2011, Article ID 416963, 12 pages, 2011.

[8] P. E. Forssén and D. G. Lowe, "Shape descriptors for maximally stable extremal regions," in *Proceedings of the 11th IEEE International Conference on Computer Vision (ICCV'07)*, pp. 1–8, Rio de Janeiro, Brazil, October 2007.

[9] C. Cattani, S. Y. Chen, and G. Aldashev, "Information and modeling in complexity," *Mathematical Problems in Engineering*, vol. 2012, Article ID 868413, 4 pages, 2012.

[10] S. Y. Chen, H. Tong, and C. Cattani, "Markov models for image labeling," *Mathematical Problems in Engineering*, vol. 2012, Article ID 814356, 18 pages, 2012.

[11] M. Li, C. Cattani, and S. Y. Chen, "Viewing sea level by a one-dimensional random function with long memory," *Mathematical Problems in Engineering*, vol. 2011, Article ID 654284, 13 pages, 2011.

[12] M. Perdoch, J. Matas, and S. Obdrzalek, "Stable affine frames on isophotes," in *Proceedings of the 11th IEEE International Conference on Computer Vision (ICCV'07)*, pp. 1–8, Rio de Janeiro, Brazil, October 2007.

[13] T. Kadir and M. Brady, "Saliency, scale and image description," *International Journal of Computer Vision*, vol. 45, no. 2, pp. 83–105, 2001.

[14] T. Tuytelaars and K. Mikolajczyk, "Local invariant feature detectors: a survey," *Foundations and Trends in Computer Graphics and Vision*, vol. 3, no. 3, pp. 177–280, 2007.

[15] D. Nistér H Stewénius, "Linear time maximally stable extremal regions," in *Proceedings of the Computer Vision-ECCV 2008*, vol. 5303, pp. 183–196, 2008.

[16] P. E. Forssén, "Maximally stable colour regions for recognition and matching," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'07)*, pp. 1–8, Minneapolis, Minn, USA, June 2007.

[17] T. Lindeberg, "Feature detection with automatic scale selection," *International Journal of Computer Vision*, vol. 30, no. 2, pp. 79–116, 1998.

[18] http://www.vlfeat.org/.

[19] http://www.robots.ox.ac.uk/~vgg/research/affine/index.html.

[20] S. Y. Chen and Y. F. Li, "Determination of stripe edge blurring for depth sensing," *IEEE Sensors Journal*, vol. 11, no. 2, pp. 389–390, 2011.

[21] H. Liu, S. Chen, and N. Kubota, "Special section on intelligent video systems and analytics," *IEEE Transactions on Industrial Informatics*, vol. 8, no. 1, article 90, 2012.