A Representative Local Region Detector Based On Color-Contrast-MSER

Yang Cao[1,2], Ke Gao[1], Sheng Tang[1], Yongdong Zhang[1]
[1] Key Lab of Intelligent Information Processing of Chinese Academy of Sciences
Institute of Computing Technology, Chinese Academy of Sciences, Beijing, 100190, China
[2] University of Chinese Academy of Sciences, Beijing, 100190, China
{caoyang,kegao,ts,zhyd}@ict.ac.cn

ABSTRACT
In order to extract representative local invariant regions in textured natural images, we propose a Color-Contrast-MSER (CCM) detector with color-contrast pixel ranking, which can reduce the number of meaningless regions extracted from backgrounds. The main contributions are threefold: (1) In contrast with the original MSER[3] which adopts intensity pixel ranking, we develop a new pixel ranking mechanism based on color contrast analysis. (2) In this paper, the pixel ranking value of each pixel is defined as the color contrast between a kernel-sized window and the background. Therefore we propose an adaptive background scale selection mechanism that simulates the background color distribution as the benchmark for color contrast. (3) The experimental results demonstrate that compared with the original MSER detector[3], our Color-Contrast-MSER (CCM) detector can extract more representative local regions with competitive repeatability score at only 50% computational time and 10% memory cost.

Categories and Subject Descriptors
I.4.10 [**Image Processing and Computer Vision**]: Image Representation

General Terms
Algorithms, Design, Performance, Theory.

Keywords
Local feature detector, Color-Contrast-MSER (CCM), pixel ranking mechanism, local invariant regions.

1. INTRODUCTION
The local feature detectors in the literature can be divided into two main categories: one is based on corners like Harris and Hessian affine detector[1]. The other category is region-based detectors such as IBR[1], EBR[1], SIFT[2] and MSER[3] detector. It is proved that region-based detector is able to achieve higher accuracy than those detectors based on corners in image retrieval applications[1]. That's mainly because the core issue of region-based detectors is to extract representative local regions. MSER detector has proven to be one of the best detectors[1]. However MSER detector is far from satisfactory as it can't detect the representative regions completely shown in Fig. 1(c). And regions must have homogeneous intensity that could be detected as stable regions of MSER[3]. We find that MSER detector[3] is limited by intensity pixel ranking. Accordingly we bring in color-contrast pixel ranking to highlight the regions of homogeneous representative color distinctive from the background, which may represent some semantic concepts in a degree. The variety of appearance and placement of semantic objects in natural images is too unpredictable. Therefore, we propose an approach that evaluates the background color distribution according to local color and texture similarities as the benchmark for color contrast.

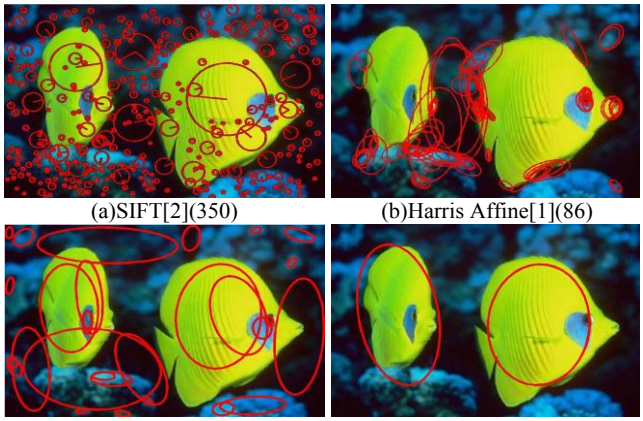

(a)SIFT[2](350) (b)Harris Affine[1](86)
(c)MSER[3](28) (d)CCM(2)

Figure 1. Sift, Harris Affine, MSER, our CCM detection results and region numbers of undersea scene.

Marta Penas et al.[4] use HSV color space and combine them into a final set of regions to enforce the distinctiveness compared to the original intensity. Aaron Chavez et al.[5] found it is reasonable in separating regions with pixel intensities from red pixels and the same operators are used for green and blue channels. The above extended approaches aim at finding regions of homogeneous color, but they neglect the effect of representativeness of regions. Michael Donoser et al.[6][7] define regions of interest and further use clustering approach to automatically identify ROIs. But the clustering is unstable and it still cannot highlight the meaningful and representative regions from the redundant and meaningless backgrounds.

In this paper, our main contributions can be concluded as threefold: (1) We propose a novel color-contrast pixel ranking

ICMR'14, April 01–04 2014, Glasgow, United Kingdom.

http://dx.doi.org/10.1145/2578726.2578782

mechanism to replace the intensity pixel ranking of MSER[3]. (2) We develop an adaptive scale selection approach to simulate background color distribution as the benchmark for color contrast. (3) Experimental results demonstrate our Color-Contrast-MSER extracts representative regions with competitive repeatability score at 50% computational time and 10% memory cost.

2. Color-Contrast-MSER (CCM)

To extract more meaningful local invariant regions, we pose color-contrast based pixel ranking to explore the regions of homogeneous representative colors. Additionally we propose an adaptive background scale selection mechanism to automatically simulate the distribution of backgrounds according to the local color and texture similarities. Then the pixel ranking is defined as color contrast calculated from the difference between a kernel-sized window and its background using KL divergence[9]. Finally homogeneous contrast region detection highlights high contrast regions from the meaningless background.

2.1 Color-Contrast Based Pixel Ranking

2.1.1 Adaptive background modeling

A basic principle in visual system is to suppress the response to frequently occuring regions, instead keeping sensitive to those regions extracted from the distinctive regions. For this perspective of visual perception theory, representative coding decomposes the image into two parts:

$$W_{image} = W_{target} + W_{background} \quad (1)$$

where W_{target} denotes the target parts and $W_{background}$ is the redundant background information.

In natural images, especially taken by photographers, it is rare to focus the targets on the corners of the image. Instead it is used to locate targets on the central or one side of the layout to fit real aesthetics of art. Statistics from [8] figure out that a 'good' image taken by a skillful photographer tends to have its ROI near the center and more than 82.2% images among 5000 web images from Google and Flickr have their ROIs located at least near the center. As such, many mechanisms directly define the central part as ROIs. However, there still exist different views shown by the images in Fig. 2, where the targets' location is not the central. Therefore we explore this problem in an alternative way: we pick up the upper left, lower left, upper right and lower right corners as the simulation of the background. Additionally we propose an adaptive scale selection mechanism of corner size to automatically imitating background content. Then we define a novel color-contrast pixel ranking that is based on analyzing the statistical characteristics of the local color and texture similarities.

Figure 2. General layout of natural images.

According to the general layout of natural images assumption, we develop an adaptive background scale selection mechanism to figure out a fitted corner size of the layout. The mechanism consists of two steps: Ⅰ. Calculation of two-dimensional Shannon entropy of local image attributes such as intensity or color over a range of scales. Ⅱ. Select scales at which the entropy over scale function exhibits a local minimum valley. The two-dimensional Shannon entropy takes the spatial information into consideration with original Shannon entropy, which represents the distribution of color with local texture similarities. The two-dimensional Shannon entropy is defined as:

$$H(s) = \sum_{i \in C} w_i H_i(s) \quad (2)$$

where C denotes the attribute channels, s denotes scales, w_i denotes the weight of the entropy and $H_i(s)$ is defined as :

$$H_R(s) = -\sum_{i \in Q} p_{ij}(c,s) \log_2 p_{ij}(c,s) \quad (3)$$

$$p_{ij}(c,s) = \frac{f(i,j)}{M \times N} \quad (4)$$

where Q denotes the quantized color-contrast levels, $P_{ij}(c,s)$ denotes the probability of f (i, j) appears in an image with the size of M×N pixels, and f (i, j) is the frequency number of pixel value i with its neighbor value j.

In step Ⅱ, scales are selected at which the entropy is a local valley. Through searching for such local valley, the most predictable content of background can be simulated. As shown in Fig. 3, from various scales of two-dimensional Shannon entropy, the entropy at scale 0.20 is a local valley. The local valley entropy indicates that most of these pixels are highly predictable and the distribution of these pixels has apparent trend.

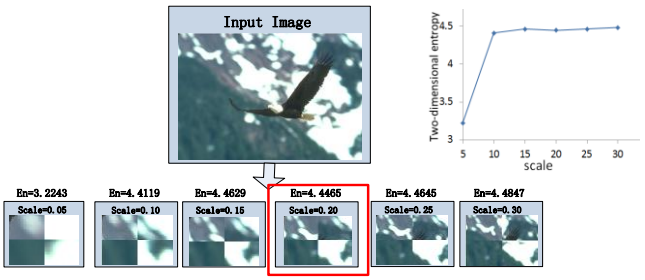


Figure 3. Adaptive Scale Selection Mechanism.

2.1.2 Pixel ranking

The original MSER detector[3] operates on intensity pixel ranking, which is not always discriminative enough as shown in Fig. 1(c). To simulate the "stand out" mechanism of visual system, the RGB color space has been experimented to be the best color space to simulate statistical characteristics in our CCM detector. We adopt KL distance[9] to measure the divergence of each pixel with its immediate surrounding areas. As single pixel is limited, a kernel-sized window is used to describe the color distribution of the central pixel, and the experimental results show that the kernel size of 15× 15 pixels is best suited. The color contrast between each kernel-sized window p_t and the adaptive background GMM distribution p_b is defined as:

$$CC(p_t, p_b) = \frac{1}{2}\{\log\frac{\sum_b}{\sum_t} + tr(\sum_b^{-1}\sum_t) - d + (u_t - u_b)^T \sum_b^{-1}(u_t - u_b)\} \quad (5)$$

where u_t, u_b denote the mean of two Gaussian distributions and Σ_t, Σ_b denote the covariance matrices. Please note that it is also possible to use JSD distance[9] instead of KL distance as well.

2.2 Homogeneous Contrast Regions Detection

MSER regions are connected regions which are converged through an iterative process. The iterative process starts with the threshold of the image intensity at all the color contrast values. At each threshold level, the pixels below the threshold are designed as black, while the pixels over the threshold are in white. An extremal property of the color function of white and black regions is a crucial decision for the MSERs as:

$$\varphi(R_i) = \frac{\left|R_j^{w-\Delta} - R_i^{w+\Delta}\right|}{R_i^w} \tag{6}$$

where $|\bullet|$ denote the cardinality, R_i^{ω} is a region that is obtained in a level set at weight ω and Δ is a stability range parameter.

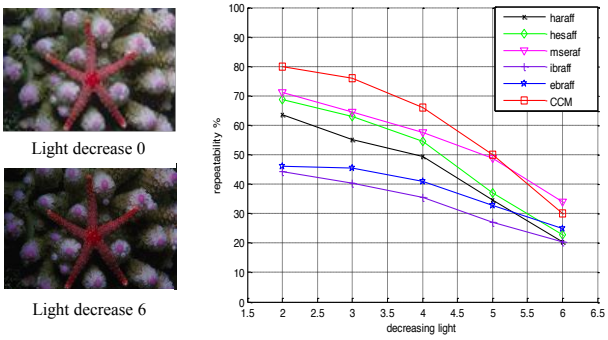
By nature our Color-Contrast-MSER (CCM) don't need to compute regions in the inverted image as Color-Contrast-MSER (CCM) refines the pixel ranking with color contrast and what we want is the high contrast representative regions shown in Fig. 1(d). Consequently compared with the original MSER detector, the time cost of each image in [10] is about 0.972 seconds running on a 3GHz CPU, our CCM only runs 0.553 seconds saving 50% computational time and only takes 10% memory cost.

3. Experiments and Analysis

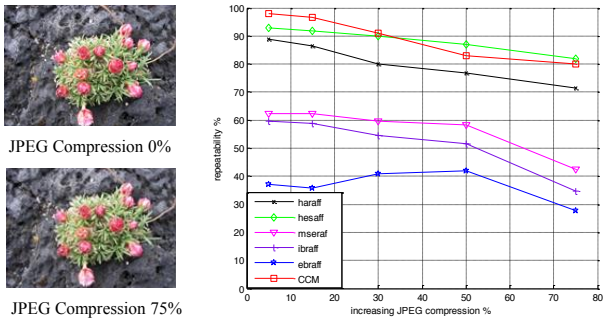
Our experiment setup is as follows. The original image passes through a scale selection module first, through which a fitted scale is selected to better simulate background color distribution. Then the pixel ranking is defined as the color contrast between a kernel-sized window and its adaptive background. Finally the bottom-up MSER only process one-time to highlight the representative regions. We give a clear comparison of our CCM detector with SIFT[2] and MSER[3] detection results on the image set[10] in Fig. 5, which shows our CCM detector actually focuses the representative regions, instead SIFT[2] and MSER[3] extract the local regions blindly and always bring in the redundant regions such as the grass, clouds and other kinds of backgrounds.

As MSER[3] has proven to be one of the best detectors in affine transformations, it still has its limitations on increasing blur and JPEG compression[1]. The experimental results demonstrate the contrastive repeatability scores of textured natural image sequences under illumination change, JPEG compression, scale change and increasing blur transformations compared with the existing affine detectors on image set[10]. In our experiments, we fix the overlap error threshold to 40% and check the repeatability of the different detectors for gradually increasing transformations according to the image sequences shown in Fig. 4.

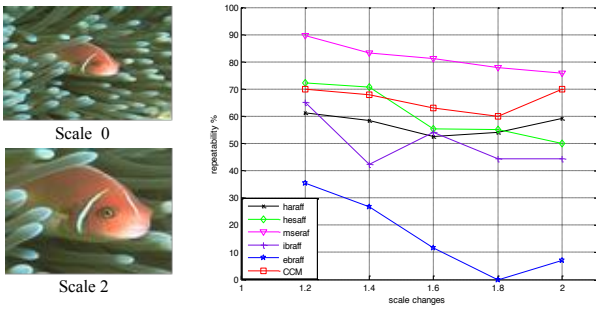
Fig. 4(a) shows the good robustness of CCM and MSER to illumination change and CCM has better performance on slight illumination change. Fig. 4(b) demonstrates the repeatability scores with increasing JPEG compression. Experimental results indicate that for this type of structured scene, with high color contrast between the visual targets and backgrounds, CCM detector is clearly best suited and the overall repeatability achieves better than the original MSER[3]. Fig 4(c) displays the repeatability score with scale changes, MSER and CCM performs best, followed by Hessian affine and Harris affine detector[1]. Fig. 4(d) shows the results for the structured scene with increasing blur. The repeatability in Fig. 4(d) for EBR detector[1] is very low as the lacking of stable edges. However, the repeatability of our CCM has better performance than MSER[3] owing to the color contrast which has better tolerance on blur than intensity alone.



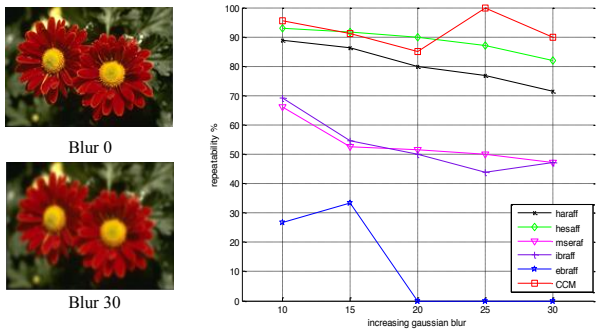
(a)Repeatability score for undersea sequence with decreasing light



(b)Repeatability score for rock flowers sequence with JPEG compression



(c)Repeatability score for undersea sequence with scale changes



(d)Repeatability score for flowers sequence with increasing gaussian blur

Figure 4. Repeatability score for illumination change, JPEG compression, scale change and increasing blur of the structured scene.

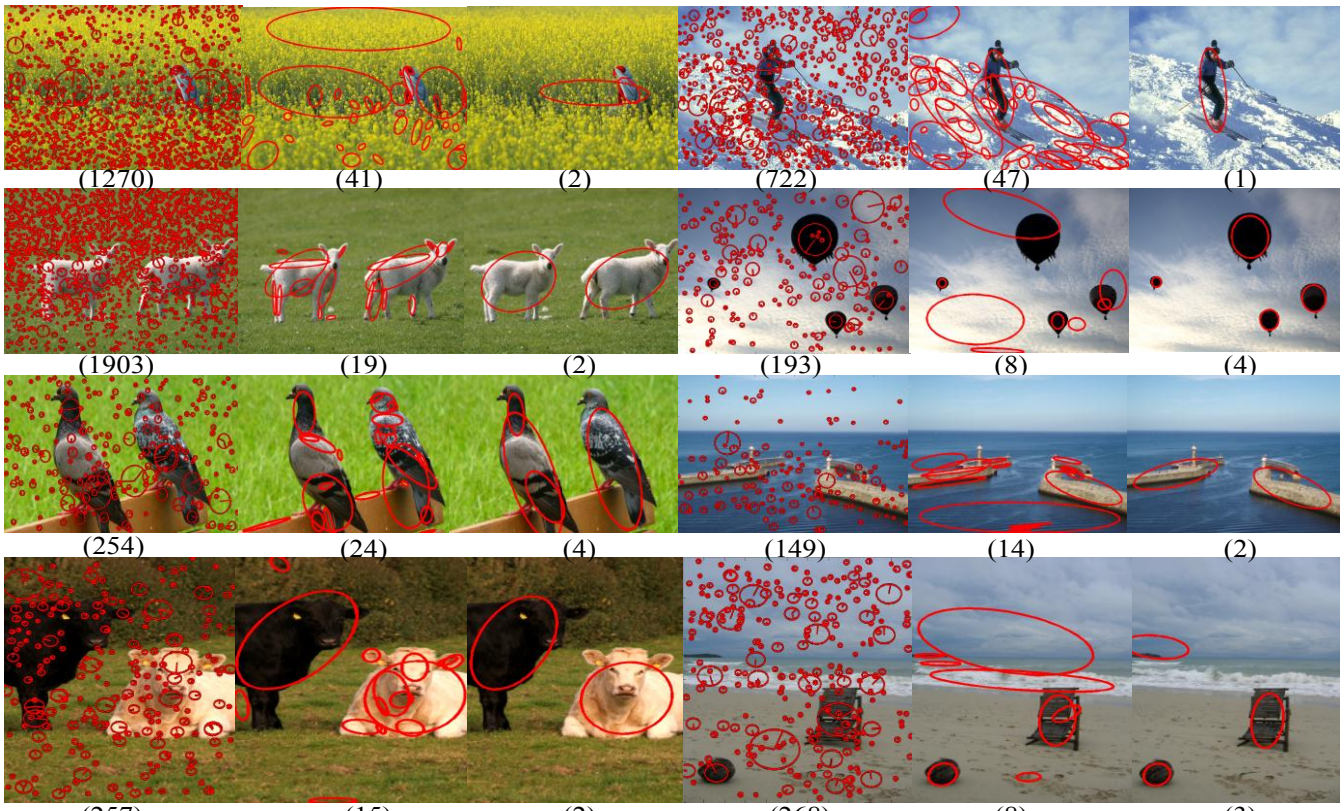


Figure 5. Some detection results and region numbers on image set[10].

4. Conclusion

This paper presents a representative local region detector based on color-contrast pixel ranking which can extract more representative local invariant regions compared with the existing local feature detectors. This color-contrast pixel ranking is based on the analysis of an adaptive background color modeling as the benchmark for color contrast. Practical experimental results show that the proposed CCM detector is competitive in repeatability at 50% computational time and 10% memory cost. The CCM detector can be easily employed in various visual content analysis related applications, such as to help object recognition and for tracking objects in videos.

5. ACKNOWLEDGMENTS

This work was supported by the National Nature Science Foundation of China (61273247, 61271428, 61173054); National High Technology and Research Development Program of China (863 Program, 2014AA015202); National Key Technology Research and Development Program of China (2012BAH39B02).

6. REFERENCES

[1] K. Mikolajczyk, T. Tuytelaars, C. Schmid, A. Zisserman, J. Matas, F. Schaffalitzky, T. Kadir, and L.V. Gool, "A Comparison of Affine Region Detectors," *International Journal of Computer Vision*, vol. 65, pp. 44-72, 2005.

[2] D.G. Lowe, "Distinctive Image Features from Scale-Invariant Keypoints," *International Journal of Computer Vision*, pp. 91-110, 2004.

[3] J. Matas, O. Chum, M. Urban, and T. Pajdla, "Robust Wide Baseline Stereo from Maximally Stable Extremal Regions," *British Machine Vision Conference*, pp. 761-767, 2002.

[4] M. Penas, and L.G. Shapiro, "A Color-Based Interest Operator," *Image Analysis and Processing*, pp. 965-974, 2009.

[5] A. Chavez, and D. Gustafson, "Color-Based Extensions to MSERs," *Advances in Visual Computing*, pp. 358-366, 2011.

[6] P.E. Forssen, "Maximally Stable Color Regions for Recognition and Matching," *CVPR*, pp.1-8, 2007.

[7] M. Donoser, and H. Bischof, "ROI-SEG: Unsupervised Color Segmentation by Combining Differently Focused Sub Results," *CVPR*, pp. 1-8, 2007.

[8] Z.Y. Chen, L.F. Sun, and S.Q. Yang, "Auto-Cut for Web Images," *the 17th ACM International Conference on Multimedia*, pp. 529-532, 2009.

[9] J.R. Hershey, and P.A. Olsen, "Approximating The Kullback Leibler Divergence Between Gaussian Mixture Models," *Speech and Signal Processing*, pp. 317-320, 2007.

[10] Z.L. Jiang, and L.S. Davis, "Submodular Salient Region Detection," *CVPR*, pp. 2043-2050, 2013.