Abstract— In this study, we propose to detect interest regions based on texture information of images. For this purpose, Maximally Stable Extremal Regions (MSER) approach is extended using the high dimensional texture features of image pixels. The regions with different textures from their vicinity are detected using agglomerative clustering successively. The proposed approach is evaluated in terms of repeatability and matching scores in an experimental setup used in the literature. It outperforms the intensity and color based detectors, especially in the images containing textured regions. It succeeds better in the transformations including viewpoint change, blurring, illumination and JPEG compression, while producing comparable results in the other transformations tested in the experiments.

Keywords- texture; maximally stable; interest region detection

I. INTRODUCTION

Detection of interest regions invariant to geometric and photometric transformations is an important topic in many computer vision applications such as object recognition, scene classification and image alignment. Various detection approaches can be divided into three groups based on the type of feature extracted in the image [1]. The first group of methods detect the corners, which correspond to the high curvature points. Harris-Affine method [2] determines an elliptical region for each corner point by utilizing the second moment matrix, which describes the local gradients. In Edge-based region detector (EBR) [3], after the corners are detected in various scales, the interest regions are detected by moving on the edges connected to the corners until some photometric quantities reach an extremum. The methods in the second group extract blobs as local features. Hessian-Affine detector [2] finds the local maxima of the determinant and trace of Hessian matrix. Scale-selection is based on Laplacian similar to Harris method. The salient region detector (SR) [4] is based on the saliency measured by the entropy of the probability distribution of intensity in an elliptical region over a range of scales around each point. Lastly, the third group of methods concern with the detection of image regions. In intensity extrema-based region detector (IBR) [3, 5], the points, where the intensity level is extremum, are used instead of the corners in EBR algorithm. An image region is marked as interest region when the rays spread out from an extremum point in all directions reach a sudden increase or decrease. Maximally Stable Extremal Regions (MSER) algorithm [6] detects the regions where the area change is local minimum over a range of intensity thresholds.

We claim that the repeatability and discriminative of a detection method can only be improved by detecting semantically meaningful interest regions. For instance, in a textured part of an image, the intensity or color based methods detect the small regions, which, in fact, construct the texture they belong. Those regions having no semantic information are sensitive to image transformations. Therefore, rather than employing pixel level features, it is necessary to use higher level features extracted in the vicinity of pixels. One solution is to develop segmentation based method such as superpixels [8]. However, the regions found by any segmentation method are relatively large and do not overlap unlike the traditional detectors. In addition, the regions may not be scale invariant and the repeatability of boundary extraction is low as indicated in [1]. Therefore, the segmentation based interest regions are less suitable for matching or object recognition.

In this study, we propose a different method utilizing texture features, which carry more discriminative information, in order to detect semantically meaningful interest regions. It extracts the connected components, which are defined as the contiguous pixels having the similar texture. For this purpose, we employ MSER approach instead of a segmentation method for increasing repeatability and locality. In the implementations, we use the framework proposed in [9] developed for maximally stable color regions. In the experiments, we compare the proposed method with the state-of-the-art detection methods using the setup in [7].

II. MAXIMALLY STABLE EXTREME REGIONS

Maximally Stable Extremal Regions (MSER) are defined as connected regions, which are consistent across several intensity thresholds. Applying binarization process with different thresholds, connected components grow and merge. The areas of all connected components, \( \zeta \), across all intensity thresholds are stored and those which have a local minimum in (1) are chosen as MSER.

\[
q(i) = \left| \frac{\zeta_{i+\Delta} \setminus \zeta_{i-\Delta}}{\zeta_i} \right|
\]  

(1)
where \(|.|\) is the area and \(\Delta\) is the user-defined margin parameter. An MSER is either darker (MSER+) or brighter (MSER-) than its surrounding.

This approach is extended to color images in [9], and named as Maximally Stable Color Regions (MSCR). It is based on successive time-steps of an agglomerative clustering of image pixels. An evolution process is defined over the image, \(I: \Omega \rightarrow \mathbb{R}^3\), where \(\Omega\) is the set of all image positions depending on the 4 or 8-connected neighborhood. The positions of neighbor pixel pairs, \(x, y\) and color distance are stored in a list. For each time step, \(\tau \in [0, T]\), a map where \(E(\tau) : \Omega \rightarrow \mathbb{R}\) is constructed. Each unique label defines a contiguous region, \(R \subseteq \Omega\). Any two pixel positions which have smaller color distance than a threshold value for that time step, \(d_{thr}(\tau)\), are put into a label image, \(E_\tau\) successively. If one of the pixel pairs belongs to a region, the other is added to that region. If both belong to different regions, the smaller region is appended to the bigger one, and if none belongs to any region, they construct new region with new label.

In MSCR, a scaled-Chi squared distribution is used to compute evolution thresholds, \(d_{thr}(\tau)\), which is similar to the intensity thresholds in MSER. For each region in the label image, the area \(a_r\) and distance threshold \(d_r\) at which the region firstly appeared are stored. When the area change, \(a_{new}\), between two time steps are more than area increment threshold, \(d_{area}(\tau)\), \(a_r\) and \(d_r\) are re-initialised. The regions which have the smallest slope, \(\sigma\), are stored. In MSCR, Ellipses are fitted to the detected regions. For this purpose, the raw moments and color moments are computed for each region represented by a binary mask. Since MSCR produces too many regions, pruning process is required for more stable color regions. Pruning process is done in three steps; the regions which have smaller margin than \(m_{min}\), the regions which have smaller area than \(a_{min}\) and the regions of which minor axis of approximating ellipse is smaller than 1.5 pixels are pruned.

III. DETECTION OF MAXIMALLY STABLE TEXTURE REGIONS (MSTR)

Our goal is to detect the regions of homogeneous texture using the maximal stability approach. The implementation of our goal is achieved by adapting the algorithm developed for detecting MSCR. In our case, a Maximally Stable Texture Region (MSTR) is defined as a connected component of the image, which is appropriately segmented by clustering neighboring pixels. The problem seems like texture segmentation. However, the clustering is applied successively in order to find the maximal stable regions. The pixels are mapped to a label at each step of an evolution process. The regions in which all the neighboring pixel distances are below a threshold are selected as MSTRs.

The first problem in the implementation is how to represent the image pixels. Each pixel should be represented with respect to the textured area around itself. This is simply achieved by the convolution of image with the filters of various scales and orientations. Each pixel is then represented by the filter responses. Another alternative is to employ only local neighborhood distribution as proposed in [10].

In our study, we tried both the popular filter banks in the literature [11, 12] and the local distributions with various neighborhood scales. One of the filter banks is proposed by Leung and Malik in [13]. It consists of multi scale, multi orientation 48 filters, including derivatives of Gaussians at 6 orientations and 3 scales, 8 Laplacian of Gaussian filters; and 4 Gaussians. Another filter bank contains rotation invariant 13 Gabor-like filters [14]. Lastly Maximum Response filter bank proposed in [11, 12] consists of 38 filters with edge and bar filters oriented at 6 orientations and 3 scales and 2 isotropic filters. The maximum response across orientations is measured and this constructs 8 responses (MR8). In the experiments, the most repeatable and discriminative regions are detected by the pixel representation using the responses of MR8 filters.

The second problem in MSTR is how to determine the steps in the evolution process for successive clustering. Note that two neighboring pixels are merged to form a region at each step, if their dissimilarity is below a threshold value, which is updated during the evolution process. The goal is to stabilize the process in such a way that the similar number of pixels is examined at each step. For this purpose, the threshold value should be updated with respect to the inverse of cumulative distribution function of distances between all the neighboring pixels. In [9], It is assumed that the edge distances follow the Chi-squared distribution. However, it is shown that spatial statistics of natural images follow Weibull type distribution in [15];

\[
f(x) = C \exp \left( -\frac{(x - \mu)^\gamma}{\beta} \right)
\]

where the parameters \(\mu\), \(\beta\) and \(\gamma\) represent the center, scale and shape of the distribution, \(C\) is the normalization constant. In our experiments we try various distribution functions, including Chi-squared and power-law, to model the distribution of distances between neighboring pixels. We find that Weibull distribution is the best fit among them. In Figure 1, the normalized histogram of the edge distances and Weibull probability density function are shown for a natural image used in the experiments.

![Figure 1. The Normalized histogram and Weibull PDF of the edge distances for the wall image.](image-url)
The parameters are estimated using the maximum likelihood method. The distance thresholds at each evolution step are then computed by the inverse Weibull distribution and stored in a look-up table. The other procedures for detecting the stable regions and restricting small unstable regions are applied as described in [9].

An example of MSTR result comparing to the other detectors for a sample textured image in our data set is given in Figure 2. Both MSER and Hessian-Affine fail in textured images due to the repeated patterns in a texture. They detect unnecessarily many regions in the Trees image. Regions detected by MSTR are both meaningfully and numerically closer to human vision system than the others, which are evaluated as the best detectors in [7].

IV. EXPERIMENTAL RESULTS

MSTR is compared with MSER, Hessian-Affine and MSCR in terms of repeatability and matching scores that measure the performance of a region detector using an image data set [7] in the literature. In the image data set, there are 8 different images that each has a reference image and 5 images transformed under geometric or photometric transformations; viewpoint change (Graffiti, Wall), scale (Boat, Bark), blur (Bikes, Trees), illumination (Leuven) and JPEG compression (UBC).

The repeatability of a detector measures how well the detector determines corresponding scene regions. Repeatability score is calculated as the ratio of the number of correspondences to the number of regions in the image which have less number of regions in both of the image pairs. The matching score measures the distinctiveness of the detected regions. It is the ratio of the number of correct matches to the number of regions in the image having fewer regions in the image pairs. After SIFT [16] descriptor vector is assigned to each region, the region pairs are assumed as correct matches if they are nearest neighbors in descriptor space.

In the experiments as shown in Figure 3, MSTR outperforms the other detectors based on maximal stability approach in viewpoint change, image blur, illumination, and JPEG compression transformations. MSTR have also higher repeatability and matching scores than Hessian-Affine method, with the exception in the UBC and Trees images. We evaluate comparable results in the other sequences of images used in the experimental setup [7]. They are not given here for lack of space. The experiments show that MSTR detects relatively large interest regions having consistent texture. We think that this is because of the pixel representation using their neighborhood instead of their grey or color value. Detecting meaningful textured regions increases its performance especially in the images containing textural information.

We also combine texture, intensity and color features as an extension to MSTR. Each pixel is represented by a feature vector constructed by the MR8 filter bank responses, grey level and \( L^*a^*b^* \) color values. However, we evaluate its performance measures are similar to those of MSTR method.

Figure 2. Interest regions. a)Trees image, b)MSER, c)Hessian-Affine, d)MSTR results in a selected image part.
Figure 3. Repeatability (left column) and matching scores (right column).
(a) Bikes (blurring), (b) Wall (viewpoint), (c) UBC (JPEG compression),
(d) Leuven (lighting), (e) Trees (blurring).

V. CONCLUSION

In this paper, we aim to develop a texture based method for detecting semantically more meaningful regions. The proposed method detects the image regions, different from their surrounding in terms of textural features. Therefore, the detected regions are more discriminative and larger than the ones detected by the available methods. We think that MSTR can be successively used in the applications, where the image is represented globally, such as scene classification. It can also be extended for texture segmentation and classification problems.

REFERENCES