Salient Region Filtering For Background Subtraction

Wasara Rodhetbhai and Paul Lewis

Intelligence, Agents, Multimedia Group,
School of Electronics and Computer Science,
University of Southampton,
Southampton, SO17 1BJ,
United Kingdom
{wr03r,phl}@ecs.soton.ac.uk

Abstract. The use of salient regions is an increasingly popular approach to image retrieval. For situations where object retrieval is required and where the foreground and background can be assumed to have different characteristics, it becomes useful to exclude salient regions which are characteristic of the background if they can be identified before matching is undertaken. This paper proposes a technique to enhance the performance of object retrieval by filtering out salient regions believed to be associated with the background area of the images. Salient regions from background only images are extracted and clustered using descriptors representing the salient regions. The clusters are then used in the retrieval process to identify salient regions likely to be part of the background in images containing object and background. Salient regions close to background clusters are pruned before matching and only the remaining salient regions are used in the retrieval. Experiments on object retrieval show that the use of salient region background filtering gives an improvement in performance when compared with the unfiltered method.

Keywords: Background Clustering, Salient Regions, Object Retrieval

1 Introduction

Salient regions are regions in an image where there is a significant variation with respect to one or several image features. In content-based image retrieval (CBIR), salient points and regions are used to represent images or parts of images using local feature descriptions. In [1, 2] the salient approach has been shown to outperform the global approach. Many researchers have proposed different techniques based on salient points and regions. For example, Schmid and Mohr [3] proposed using salient points derived from corner information as salient regions for image retrieval, whilst Q.Tian [4] *et al* used a salient point detector based on the wavelet transform.

Salient regions are also applied to the problem of object retrieval, for example, in the case where a specific object in a query image is required to be retrieved from the image database. Traditional CBIR based on salient regions begins with salient region detection. Each salient region is then typically represented by a feature vector extracted from the region. In the query step, there is matching between salient regions from the query image and those from images in the collection and similar images are ranked according to the quality of match.

However, one of the reasons that the accuracy of object retrieval may be less than optimal is the presence of salient regions in the retrieval process which are not located on the object of interest.

Attempts to reduce the influence of irrelevant regions have appeared in some research projects. Ling Shao and Michael Brady [5] classify the selected regions into four types before the use of correlations with the neighbouring region to retrieve specific objects. Hui Zhang [6] *et al* pruned salient points using segmentation as a filter.

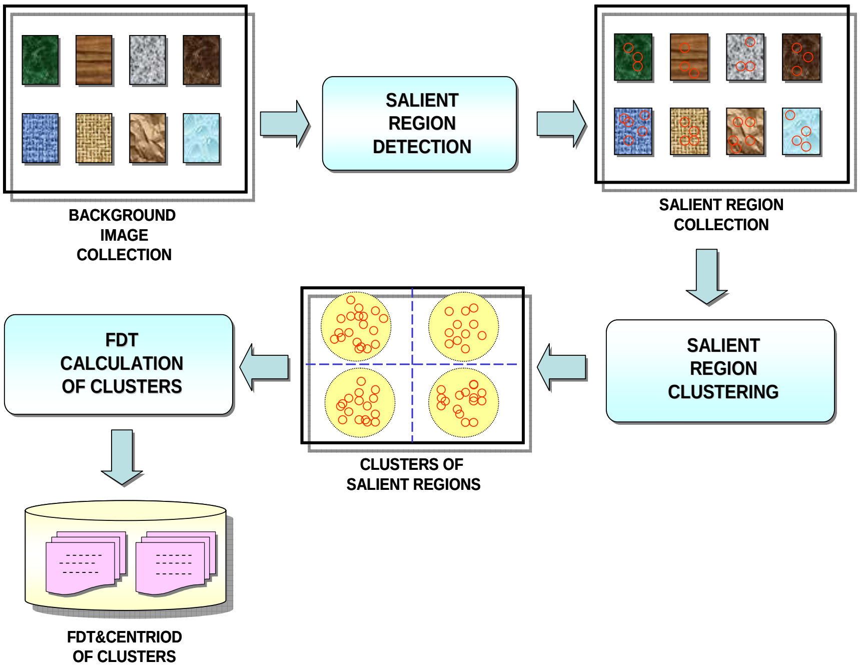

BACKGROUND IMAGE COLLECTION
SALIENT REGION DETECTION
SALIENT REGION COLLECTION
SALIENT REGION CLUSTERING
CLUSTERS OF SALIENT REGIONS
FDT CALCULATION OF CLUSTERS
FDT&CENTRIOD OF CLUSTERS

Fig. 1. Background clusters for filtering salient regions

In this paper we propose a method to filter salient regions using background information. Situations where the technique may be particularly appropriate are those where the image backgrounds are not completely arbitrary but can be characterized by a limited number of prototypes. Identifying particular objects in indoor scenes is an example.

In our approach, the system begins by creating clusters of salient regions from a collection of background only images. Thereafter, when processing images containing objects, salient regions with a high probability of belonging to a background cluster are removed before further processing. The process will be described in Section 2. It is illustrated schematically in Figure 1 and uses a distance threshold from the centre of each cluster called the fractional distance threshold (FDT).

For image retrieval, the background filtering step is applied after salient regions have been extracted. The process is illustrated schematically in Figure 2.

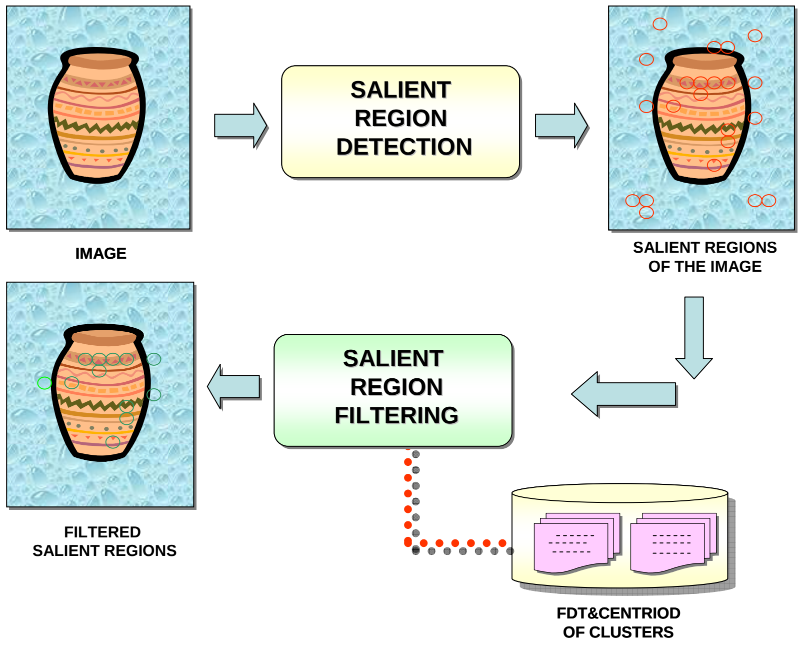


Fig. 2. The salient region filtering process

In the following sections we show that by using salient region filtering it is possible to reduce the number of unwanted salient regions and improve the precision of the retrieval process.

The paper is organised as follows. In Section 2, the methods of background clustering and calculation of the fractional distance threshold are introduced. The experimental procedure is described in Section 3. Results and discussion are presented in Section 4 and finally, Section 5 presents the conclusions and a brief discussion of future work.

2 Background Clustering

2.1 Salient Region Detection and Feature Extraction

Recently, many local detectors which can identify salient regions in an image have been described and evaluated [7, 8]. One of the popular approaches to salient region detection and representation is to use the multi-scale difference-of-Gaussian (DoG) pyramid for region location and scale estimation and the SIFT (Scale Invariant Feature Transform) from Lowe [9] to represent the detected salient regions. For each salient region, a 3D histogram of gradient locations and orientations is calculated. The SIFT descriptor has been evaluated by Mikolajczyk and Schmid in [10] to be one of the best performing local descriptors. The DoG and the SIFT approaches to salient region detection and representation are those adopted in our work.

2.2 Background Cluster Construction

One assumption of the method presented here is that salient regions from the foreground objects are reasonably distinct from background salient regions, or that any similarities involve a sufficiently small proportion of the total object salient regions to make their removal negligible.

The method begins with the detection of salient regions in a collection of background images. Since large numbers of salient regions may typically be detected in a single image, a random sample of salient regions are selected from all the background images and feature descriptors are extracted and used to cluster the salient regions into k clusters using the k-means clustering algorithm.

Since the clusters are derived from salient regions on background only images, these clusters are identified as the *background clusters*. The centroid of each cluster is calculated, essentially as a 128 element SIFT descriptor. Members in the same cluster are background salient regions that are similar to each other and dissimilar to the salient regions of other groups.

Many of the clusters are quite small in number so deriving a valid statistical model of the background clusters was not possible but to discriminate between salient regions on foreground and background, we determine an appropriate percentile distance from each cluster centroid, which we call the Fractional Distance Threshold (FDT) for each of the background clusters. The FDT of a cluster is the distance between a cluster member at a particular percentile and the centroid of that cluster. Thus FDT (90) is the distance from the centroid to a cluster member for which 90% of cluster members are nearer the centroid. The same percentile is used for all clusters and the appropriate percentile value found by experiment (see Section 3). The actual FDT and centroid for each cluster is retained for use in the retrieval process.

In the salient region filtering step, salient regions are detected and the features extracted. Any salient region (*S*) which has a feature distance (*D*) to the centroid (*C*) greater than the FDT (*i*) value for all (*n*) clusters, is assumed to be a salient region on the foreground (S_F). Otherwise, it is assumed to belong to a background region (S_B) as is represented by the following formula.

$$S = \begin{cases} S_F & \text{, if } \forall \mathrm{k} \left[\mathrm{D}(\mathrm{S}, \mathrm{C}_\mathrm{k}) > \mathrm{FDT}(\mathrm{i})_k \right] \\ S_B & \text{, otherwise} \end{cases} \quad \textbf{(1)}$$

$$\text{where } \mathrm{k} \in \{1, 2, ..., \mathrm{n}\} \text{ and } \mathrm{i} \in \{1, 2, ..., 100\}$$

3. Experiment

We separate the experimentation into 2 parts. The first part is to establish appropriate parameters for the clustering and Fractional Distance Threshold estimation and the second is to evaluate the retrieval performance using background salient region filtering.

3.1 FDT Percentile Estimation

A background only image collection, composed of 120 background only images (400 x 300 pixels) was created for 12 different backgrounds (10 images per background). Salient regions were extracted from each of the images and the number of salient regions found in each image varied between 8 and 3,503 depending on image content. Figure 3 shows some example background images from the dataset.

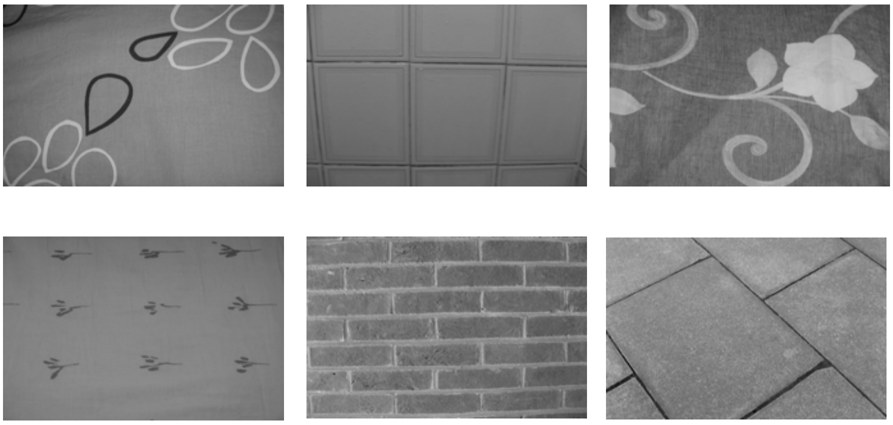
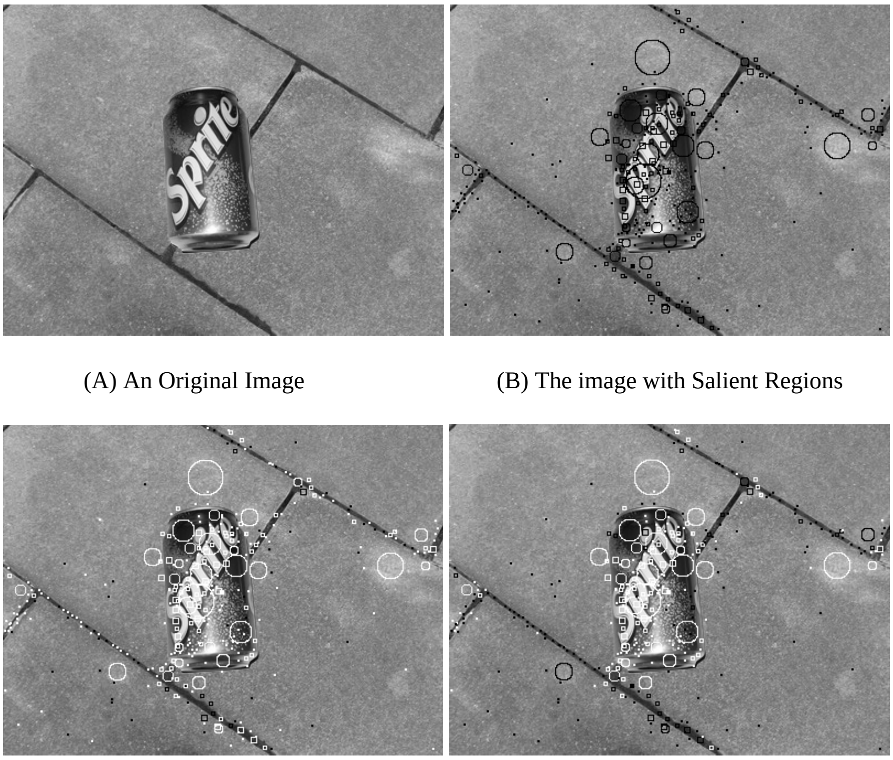


Fig. 3. Sample background images

In order to find appropriate values for the number of clusters in the k-means clustering, *k*, the number of randomly selected salient regions to use, *S*, and the percentile setting for the FDT calculation, a range of *k* and *S* combinations was used for clustering and the FDT estimated at each percentile from 50 to 100 in steps of 5. Each of eleven different *k* and *S* combinations were used. The resulting FDTs were used to check the percentage of correctly assigned foreground and background salient regions on a collection of object and background images. For these images, the ground truth was established by manually delineating the area covered by the object and if the centre of a salient region (SR) fell in the object area it was taken as an object SR. Otherwise, it was taken as a background SR.

Figure 4 shows examples of the decisions made by the system using some of the different FDT values. Salient regions in white circles represent foreground SRs and those with black circles represent background SRs.



(A) An Original Image

(B) The image with Salient Regions

(C) The image with filter at FDT = 50

(D) The image with filter at FDT = 80

Fig. 4. Foreground (white circle) and Background (black circle) salient regions at FDT = 50 and FDT = 80

The performance of the decisions for a range of *k*, *S* and FDT values is measured via the receiver operating characteristic (ROC) space [11]. A ROC space represents the relationship of true positive rates (TP) and false positive rates (FP). Each classification produces a (TP and FP) pair corresponding to a single point in ROC space. We define

w as the number of **correct** predictions that an instance is **Foreground SR**
x as the number of **incorrect** predictions that an instance is **Background SR**
y as the number of **incorrect** predictions that an instance is **Foreground SR**
z as the number of **correct** predictions that an instance is **Background SR**

The recall or true positive rate (TP) determines the proportion of background SRs that were correctly identified, as calculated using the equation:

$$TP = \frac{z}{y+z} \tag{2}$$

The false positive rate (FP) defines the proportion of foreground SRs that were incorrectly classified as background SRs, as calculated using the equation:

$$FP = \frac{x}{w+x} \tag{3}$$

Figure 5 shows the ROC curve (TP against FP) as the FDT percentile is varied from 50 to 100.

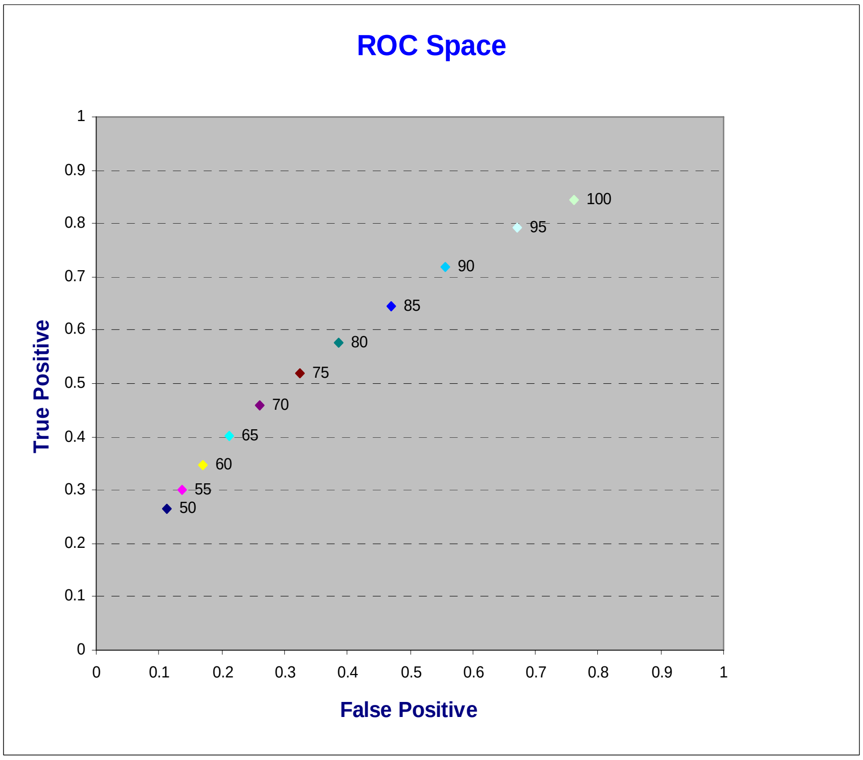


Fig. 5. The TP and FP coordinates of DFT at 50 to 100 on the ROC space

To comparing the prediction performance, distances are calculated from all points to the perfect classifier point in ROC space which is the point (0, 1). The point (0, 1) means all regions are classified correctly.

The overall results are presented in Table 1 where, for each of the *k* and *S* combinations, the table shows the distance from all of the TP and FP pairs to the point (0, 1). It illustrates how the percentage correct varies with the percentile for the FDT. It can be seen that in general a percentile of 85 gives the best results and that this is achieved with a *k* value of 5,000 and an *S* value of 50,000. These values were used in the retrieval experiments in the following section.

Table 1. The distance to (0,1) of 11 background cluster types (A-K) at the different FDT value (50 – 100). The lower the distance, the better the classifier.

Background Cluster	FDT										
(k - cluster, S - sample)	50	55	60	65	70	75	80	85	90	95	100
A (k500,S5000)	0.7528	0.7332	0.6997	0.6661	0.6336	0.6134	0.6038	0.6121	0.6408	0.7325	0.7893
B (k500,S10000)	0.6747	0.6460	0.6104	0.5874	0.5647	0.5675	0.5987	0.6381	0.6976	0.7547	0.8831
C (k500,S50000)	0.6048	0.5725	0.5474	0.5469	0.5600	0.5880	0.6332	0.7012	0.7755	0.8759	0.9778
D (k500,S100000)	0.5925	0.5614	0.5388	0.5335	0.5488	0.5809	0.6403	0.7145	0.7951	0.8967	0.9859
E (k1000,S5000)	0.8536	0.8431	0.8155	0.7786	0.7204	0.6799	0.6574	0.6285	0.5791	0.6279	0.6334
F (k1000,S10000)	0.7871	0.7635	0.7200	0.6646	0.6253	0.5931	0.5669	0.5840	0.5968	0.6954	0.7600
G (k1000,S50000)	0.6637	0.6198	0.5797	0.5466	0.5305	0.5456	0.5691	0.6295	0.7074	0.8094	0.9363
H (k1000,S100000)	0.6494	0.6040	0.5645	0.5315	0.5200	0.5374	0.5799	0.6621	0.7448	0.8612	0.9658
I (k5000,S10000)	0.9731	0.9730	0.9687	0.9556	0.9473	0.9437	0.9372	0.8842	0.7693	0.7304	0.7291
J (k5000,S50000)	0.8637	0.8439	0.7962	0.7437	0.6926	0.6214	0.5667	0.5149	0.5304	0.6075	0.6864
K (k5000,S100000)	0.8206	0.7886	0.7470	0.6902	0.6254	0.5733	0.5179	0.5194	0.5610	0.6440	0.7763

3.2 Object Retrieval

In order to test the effectiveness of background filtering, two datasets, each of 120 individual object images, were created from 10 objects on 12 different backgrounds which are not duplicated from the background training dataset. In dataset 1, the number of salient regions in these images varied between 98 and 1,728 regions. There are no scale and orientation change in each object. In dataset 2, the number of salient regions per image is between 174 and 2,349 regions. The scale and orientation is varied. From the results of the clustering experiments described earlier, the 5,000 background clusters from 50,000 salient points were used as the background clusters in the retrieval experiment and the chosen FDT value for all clusters was set to 85. Each object image was used in turn as the query image. The salient regions were extracted and background salient regions were filtered out from both the query image and the remaining object dataset images. After pruning, the strongest 50 salient regions from the remaining SRs were used to calculate the similarity between the query and dataset images and precision and recall results were obtained. The experiment was repeated without background filtering.

4 Results and Discussion

The precision and recall graphs with and without background filtering are shown in Figure 6 for dataset 1. From the graph it can be seen that the object retrieval system with background filtering outperforms the system without background filtering with an improvement in precision. The average precision with background filtering is 0.2483 and without background filtering average precision is 0.1810.

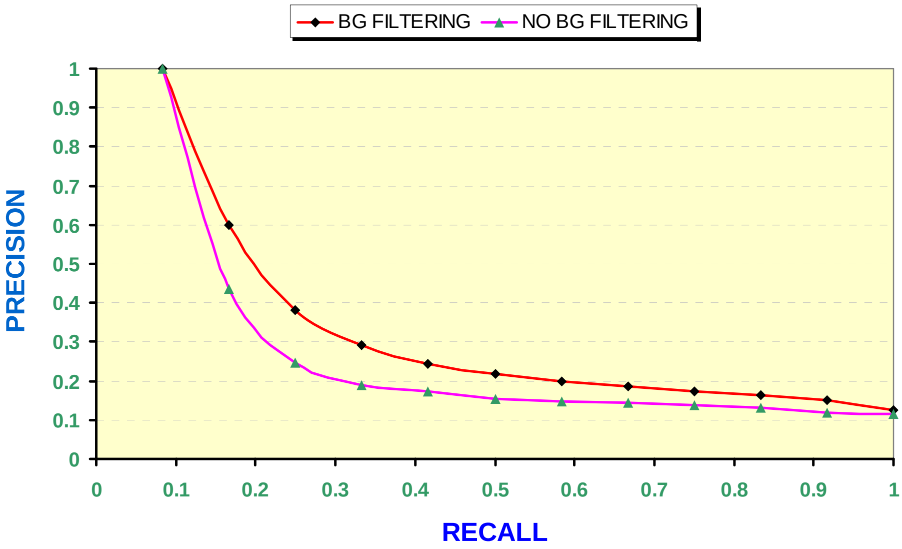


Fig. 6. Dataset 1. Precision and recall with and without background filtering

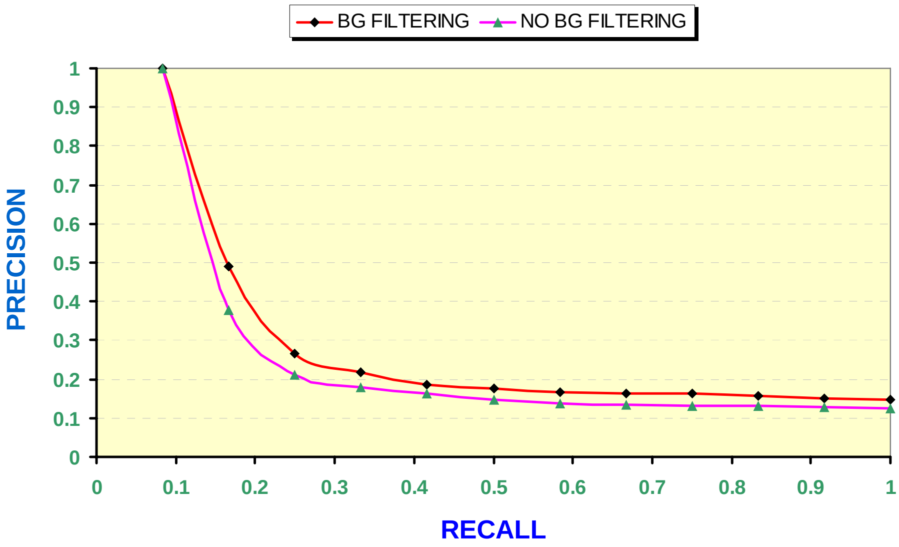


Fig. 7. Dataset 2. Precision and recall with and without background filtering

For the more challenging dataset 2, the precision and recall graph is shown in Figure 7. Again the performance is improved by using background filtering. The average precision is 0.2393 without background subtraction and is 0.2740 with background subtraction.

Looking back to Figure 5 it can be seen that best salient region classification performance was still far from the perfect classifier. The salient region filtering uses the particular differences between object areas and background areas to discriminate these regions and this is clearly not very robust.

5 Conclusion and Future Works

A novel method of filtering background salient regions for object retrieval is developed and implemented. A comparison has been made between retrieval with and without background salient region filtering and the filtering process is found to give improvements in precision. This was a rather preliminary evaluation of the technique and a more substantial evaluation is planned together with a search for a more robust way of modeling the backgrounds in terms of salient regions.

The main future work is developing a method for discriminating effectively between the object and background salient regions. More powerful feature descriptors will be incorporated to represent salient regions once identified. For the current method, more evaluation on the scale and rotation image dataset is required.

Another way to improve the performance of salient region filtering is to introduce techniques for modeling the object classes, in cases where these are known, rather than or in addition to the modeling of the background.

In summary, the background filtering method is an attempt to distinguish between the objects and the surrounding areas. Since certain types of query can benefit from using background information to filter irrelevant regions further attempts are being made to improve performance of this technique.

Acknowledgments. Our thanks to Jonathon S Hare for assistance with the salient region extraction process system.

References

1. Sebe, N., Tian, Q., Loupias, E., Lew, M.S., Huang, T.S.: Evaluation of Salient point Techniques. In: Proc. of International Conference on Image and Video Retrieval (CIVR'02), London, UK (2002) 367-377
2. Hare, J.S., Lewis, P.H.: Salient Regions for Query by Image Content. In: Proc. of The Thrid International Conference on Image and Video Retrieval (CIVR'04), Dublin, Ireland (2004) 317-325.
3. Schmid, C., Mohr, R.: Local Grayvalue Invariants for Image Retrieval. IEEE Transactions on Pattern Analysis & Machine Intelligence **19** (1997) 530-535

4. Tian, Q., Sebe, N., Lew, M.S., Loupias, E., Huang, T.S.: Image Retrieval using Wavelet-based Salient Points. Journal of Electronic Imaging, Special Issue on Storage and Retrieval of Digital Media **10** (2001) 835-849
5. Ling Shao, M.B.: Specific Object Retrieval Based on Salient Regions. Pattern Recognition **39** (2006) 1932-1948
6. Zhang, H., Rahmani, R., Cholleti, S.R., Goldman, S.A.: Local Image Representations Using Pruned Salient Points with Applications to CBIR. In: Proc. of the 14th Annual ACM International Conference on Multimedia (ACM Multimedia) (2006)
7. Fraundorfer, F., Bischof, H.: A Novel Performance Evaluation Method of Local Detectors on Non-planar Scenes. In: Proc. of Computer Vision and Pattern Recognition (CVPR) (2005) 33-33
8. Moreels, P., Perona, P.: Evaluation of Features Detectors and Descriptors Based on 3D Objects. In: Proc. of 10th IEEE International Conference on Computer Vision (ICCV 2005) (2005) 800-807
9. Lowe, D.G.: Distinctive Image Features from Scale-Invariant Keypoints. International Journal of Computer Vision **60** (2004) 91-110
10. Mikolajczyk, K., Schmid, C.: A Performance Evaluation of Local Descriptors. IEEE Transactions on Pattern Analysis & Machine Intelligence **27** (2005) 1615-1630
11. Fawcett, T.: ROC Graphs: Notes and Practical Considerations for Researchers. Machine Learning (2004)