Segmentation and region of interest based image retrieval in low depth of field observations

Rajashekhara a,*,1, Subhasis Chaudhuri b

a Global Diagnostic X-ray Engineering, GE Healthcare Technologies, Bangalore 560066, India
b Vision and Image Processing Laboratory, Department of Electrical Engineering, Indian Institute of Technology-Bombay, Mumbai 400076, India

Abstract

In this paper we address the problem of extracting the focused region and its use in retrieving similar images from a low depth of field image database. We compute the histogram of the local contrast at each pixel and model it as a mixture of two exponential distributions – one for the focused and the other for the defocused region. Unlike the mixture of Gaussian distributions, a mixture of exponential distributions overlaps with same monotonicity over the entire range in \([0, 1]\) and it is difficult to separate into components. We estimate the parameters of these distributions using the EM algorithm. This is followed by a hypothesis testing which segments the focused region in the low depth of field image. A content-based retrieval scheme is now confined to the detected region for a proper retrieval. Experimental results for both segmentation and image retrieval using a database consisting of 4986 images are presented to show the efficacy of the suggested scheme.

© 2007 Elsevier B.V. All rights reserved.

Keywords: Contrast; EM algorithm; Low depth of field; ROI; Color histogram; Precision; Recall

1. Introduction

In various application domains such as entertainment, education, biomedicine, and crime prevention, the volume of multimedia data archives is growing rapidly. The very large repository of digital information raises challenging research problems in content-based image indexing and retrieval (CBIR). One of the earliest works on CBIR includes query-by pictorial example [1]. Niblack et al. [2] developed a first general purpose content based image retrieval system QBIC nearly ten years ago. A number of general purpose CBIR systems [3,4] have been built since then which are capable of supporting new algorithms as they are developed. Recent studies have highlighted the fact that features like color, texture, shape, and spatial position and object motion etc., indeed possess a very high semantic value, and are effectively used in several CBIR systems proposed in [5–8].

Color is one of the most extensively used visual content for image retrieval. Shettini et al. [9] and Del Bimbo [6] provide a comprehensive survey of various methods employed for color image indexing and retrieval in image databases. Color histograms are useful because they are relatively insensitive to position and orientation changes. It is very easy to compute and is effective in characterizing both the global and local distribution of colors in an image. One such early work on this was proposed by Swain et al. in [10]. Considering that some of the histograms are very sparse and thus sensitive to noise, authors proposed in [11] the cumulative color histogram. To overcome the quantization effects they subsequently presented an approach based on color moments. The first three moments have been shown to be efficient and effective in representing color distributions of images. In [12], to incorporate spatial information into the color histogram, the use
of color coherence vectors has been proposed. Each histogram bin is partitioned into two types i.e., coherent, if it belongs to large uniformly-colored region, or incoherent, if it does not. The color correlogram [13] has been proposed to characterize not only the color distribution of pixels, but also the spatial correlation of pairs of colors. Based on a multiresolution framework, authors proposed in [14] a retrieval technique using hierarchical histogram. Jhanwar et al. have presented in [15] a translation and illumination invariant retrieval scheme using motif co-occurrence matrix. It uses an optimal Peano scan to encode the image. Since color histogram does not capture the shape information, a scheme based on the color histogram in conjunction with edge information has been proposed in [16].

One recent work proposed in [17] on image retrieval is based on edge feature extraction. Using the edge map in the image, a fuzzy compactness feature vector is computed which is subsequently used for similarity measurement.

Many researchers have also exploited the use of regions during retrieval. A region based CBIR system called WINDSURF has been proposed in [18]. It uses discrete Wavelet transforms to extract a set of features representing each image in the color-texture space. These features are subsequently used to partition the image into a set of homogeneous regions. Finally, Bhattacharyya distance is used to compute the similarity between images. The work of Wang et al. in [19] involves searching in high resolution pathology images for blob type of objects which possibly represent an abnormality. The SIMPLIcity system proposed in [20] uses semantics, wavelet-based features, and region matching. Regions are characterized by color, texture, shape, and location. Lipson et al. [21] also retrieve images based on spatial and photometric relationships within and across simple image regions. In [22], authors perform retrieval based on the segmented image regions. However, the segmentation algorithm includes an optional manual region pruning step, and the user must specify the expected number of regions making it unsuitable for a wider acceptance.

Our approach is, in some sense, is similar to Carson et al. [23] who perform retrieval based on segmented image regions. Here, the segmentation algorithm uses an EM algorithm to estimate the parameters of a mixture of exponential distributions of the local contrast features in a low DOF observation.

In all work on CBIR, it has always been assumed that the images are captured with a pin-hole camera, i.e., there is no depth related defocus blurring. Unfortunately all commercially available cameras are of real-aperture and whenever there is a depth variation in the scene, it will introduce a defocus blur in the image [25]. All images meant for machine vision purposes are usually captured with a very low aperture size, and under a controlled environment, when one need not worry about the defocus blur. However, such images are mostly the dull one from the human perspective and have no artistic value. Artists prefer capturing their subjects (read beautiful models or nature) with a soft focus, i.e., with a large aperture. It is the soft focus that add artistic value to a given composition. A quick look at the National Geography magazine or some glamour magazines will attest to the above fact. Hence, in spite of photography being the profession, most artists cannot use existing CBIR systems. Further, current CBIR methods cannot be applied to the photo-coverage of small and large area games like soccer, cricket, basket ball, etc., due to a large disparity in depth among players and spectators. Due to the defocus blur, the image features change quite drastically and the CBIR method yields very poor results. This is the motivation for developing a CBIR method tuned to low depth of field (DOF) images. Unfortunately this involves solving a very difficult problem of having to segment the image into focused and defocused regions, i.e., we must know the layers of depth in the scene. Layer based segmentation is quite common in a video sequence as one can use motion cue [26]. One can solve the layer based segmentation using defocus cue if one is provided with multiple observations [25]. Our problem is to perform the segmentation using only a single image! In this paper we show how this can be achieved.

This paper focuses on developing a framework for extracting the focused region and hence retrieve similar low DOF images based on the features from the focused region. For convenience, we call the sharply focused entities of such a low DOF image as the region of interest (ROI) and everything else as the background. We initially compute the histogram of the local contrast at each pixel and model it as a mixture of two exponential distributions – one for the focused and the other for the defocused region. We then estimate the parameters of the two functions using the EM algorithm. Using the Bayes theorem we compute the posterior probabilities at each pixel of the low DOF image. This subsequently follows a hypothesis testing which extracts the focused region in the low DOF image. Finally we use a CBIR confined only to the extracted focused region. The proposed technique being a region based one, it will have similar advantages or disadvantages like Blobworld and SIMPLIcity CBIR methods.

We assume in this paper that focused region in the scene always corresponds to the region of interest (ROI). Further, we assume that there is no significant variation in
depth in the ROI so that no part of ROI is defocused, leading to partial segmentation. This is tantamount to assuming that the ROI is relatively flat compared to its background. However, we make no assumption on how many distinct objects are there in the ROI. Neither do we make any assumption on their spatial contiguosity.

The remainder of the paper is organized as follows: In the following section we discuss the contrast feature and the use of its distribution. In Section 3 we discuss the issue of ROI segmentation. This section describes the mixture model for the contrast distribution, estimation of the model parameters, and quantitative analysis of the pixel classification issues. Section 4 explains the choice of feature selection for doing CBIR. This is followed by discussion on experimental results in Section 5. Finally, we conclude in Section 6.

2. Local contrast as a feature

The definition of the local contrast for an image point at the position \( c(i,j) \) is the ratio between the absolute value of the two dimensional gradient of intensity \( (\nabla I) \) and local mean intensity \( I_{mean}(i,j) \), i.e., \( c(i,j) = \frac{|I(i,j)|}{I_{mean}(i,j)} \). Here, the mean is taken over the four immediate neighbors, and the gradient is computed using first order partial derivatives. We use this definition of local contrast, since it is the most relevant psychovisual variable for edge localization and perception. Balboa et al. [27] have studied the properties of the distribution of the local contrast in great details under many different types of imaging systems, such as, outdoor picture, underwater, and atmospheric disturbances. They found that the contrast histogram, in general, has an exponential distribution \( f_c(x) = \lambda e^{-\lambda x}U(x) \). For a good quality outdoor picture \( \lambda \) is quite small and hence there are large number of pixels in the image having a large contrast value. Here the image looks very crisp and offers a high level of clarity. However, if the image passes through a scattering media such as in underwater photography, or if the image is quite blurred, the local contrast in the image decreases. Balboa et al. found that the contrast distribution was still exponential but with a large value of \( \lambda \). These issues of the contrast distributions inspired us to exploit its use in modeling the focused and the defocused regions of the low DOF image.

In low DOF image the distribution of contrast is derived from both the sharply focused part of the scene and from its defocused region. Due to defocus blur, the contrast feature of the defocused region change quite drastically from that of the focused region. We illustrate this with examples. In Figs. 1(a–d) show four different low DOF images. Notice that the background in all cases is highly defocused. For illustration, we manually crop out ROI (focused) regions in all these images and plot the corresponding contrast distributions (histograms) for ROI and the background regions in a semilog scale in Figs. 2(a–d), respectively. Notice that up to knee region, all these curves can, indeed, be approximated by a linear function, reinforcing the conjecture of Balboa et al. [27]. Further we observe that the slope for the ROI case is less steep than that of the background region.

3. ROI segmentation

3.1. Mixture model for contrast distribution

In vision and pattern recognition community, most of the researchers have invariably used Gaussian distributions as constituents while using a mixture model for various applications [23,28]. As explained earlier, we observe that contrast distributions over both the focused and defocused regions can be modeled as separate exponential functions. Since ROI region in the image occupies an arbitrary fraction of the data, the overall contrast distributions can be given as a mixture of exponential distributions.

\[
 f_c(x) = f(x|F)p(F) + f(x|DF)p(DF),
\]

\[
 = \sum_{j=1}^{2} \pi_j \exp(-x/\mu_j)U(x). \tag{1}
\]

Where \( x \geq 0 \) denotes a particular contrast level, F and DF correspond to focused and defocused regions, respectively, \( \pi_j/\mu_j \) are the mixing proportions (0 \( \leq \pi_j \leq 1 \)), \( \sum_{j=1}^{2} \pi_j = 1 \), and \( \mu_j \) denotes the mean contrast of the \( j \)th class.

Typically \( \mu_1 > \mu_2 \) as \( \mu_1 \) corresponds to the contrast distribution in the focused region. It is very difficult to separate a mixture of exponential distributions, unlike in case of a mixture of Gaussian distributions with different means. This is due to the fact that both the components overlap (monotonically) over the entire contrast range \( c \in [0, \infty) \). This is illustrated in Fig. 3. Fig. 3(a) shows a mixture of Gaussian distributions where an optimal threshold \( T \) can be found for which the probability of misclassification is quite low. For a mixture of exponential distributions, even for the optimal threshold of \( T \) as shown in Fig. 3(b), the probability of misclassification will be still quite high even when the \( \lambda \)'s are well separated. Further, since the exponential distribution has a longer tail, the separation of mixing elements is very difficult.

It may be noted that we implicitly assume that the objects in the scene are primarily at two widely separated depth levels, one of which is in focus and while the other is in very blurred. In case the scene has continuous depth variation, one needs a larger number of components in the mixture distribution estimation of which would be even more difficult. We make use of the EM algorithm to estimate the mixture distribution which will be discussed in the following section.

3.2. Estimation of model parameters

In the pattern recognition community many researchers have used the EM framework [23,29] for parameter estimation. The advantage of using the EM algorithm [30] is that it is very easily programmable and it satisfies a monotonic convergence property. If there are multiple maxima, some-
times computation of the global maximum may be erroneous and it mainly depends on the choice of good initial values. In order to estimate the parameters of the distributions of the focused and defocused regions we exploit the use of EM technique in our approach.

The EM algorithm is used for maximum likelihood parameter estimation when there is missing or incomplete data. In our case the missing data is the cluster to which the points in the feature space belong. This is computed by estimating the parameters \((\pi_1, \pi_2)\) and \((\mu_1, \mu_2)\) in Eq. (1).

The first step in applying the EM algorithm is to initialize the parameters \(\mu_1, \mu_2\) and \(\pi_1, \pi_2\). We initially start with some random values. We then use the following update equations:

**Step1 (E – step)** Given the current values \(\mu_j^{\text{old}}\) and \(\pi_j^{\text{old}}\), \(j = 1, 2\) we calculate the probability \(w_j\) that the observation \(x_i\) belongs to the \(j\)th region after observing it, i.e. the probability of \(x_i\) belonging in the \(j\)th region

\[
W_j = \frac{\pi_j^{\text{old}} f(x_i | \mu_j^{\text{old}})}{\sum_{j=1}^{10} \pi_j^{\text{old}} f(x_i | \mu_j^{\text{old}})}.
\]

**Step2 (M – step)** Update the estimates as
Fig. 3. Illustration of difficulty in hypothesis testing for exponentially distributed data. The shaded regions show classification errors for (a) Gaussian distributions, and (b) exponential distributions. The local contrast is plotted along the horizontal axis. Shaded regions correspond to classification errors.

\[ p_j^{\text{new}} = \frac{\sum_{i=1}^{n} w_{ij} x_i}{\sum_{j=1}^{p} w_{ij}}, \]
\[ n_j^{\text{new}} = \frac{\sum_{i=1}^{n} w_{ij}}{n}, \]

where \( n \) is the number of contrast levels \( x_i \).

**Step 3** Check if some stopping condition is satisfied in order to terminate the iterations, otherwise go back to step 1, for more iterations.

Our stopping criterion is based on the relative change of parameters over successive iterations. The maximum over all the parameters is used as the criterion, so the criterion takes the form

\[ \max \left\{ \frac{\left| \mu_1^{t+1} - \mu_1^t \right|}{\mu_1^t}, \frac{\left| \mu_2^{t+1} - \mu_2^t \right|}{\mu_2^t}, \frac{\left| \pi_1^{t+1} - \pi_1^t \right|}{\pi_1^t}, \frac{\left| \pi_2^{t+1} - \pi_2^t \right|}{\pi_2^t} \right\} < \epsilon \]

where \( \epsilon \) is a small number (we select \( \epsilon = 10^{-4} \)). Experimentation shows (see Fig. 6) that the parameters converge after repeating the steps 1 and 2 for 1000 iterations and finally satisfy the stopping criterion. The parameters thus computed describe the conditional contrast distributions of the focused and the defocused components which are subsequently used to compute the posterior probabilities of the pixels for grouping them into one of the two regions.

### 3.3. Pixel labeling

In order to classify pixels into ROI or background classes we deploy the principle of Bayesian decision theory. The parameters estimated in Section 3.2 are used to compute the posterior probabilities of belonging to a given class for all pixels in the low DOF image. The posterior probability for each pixel belonging to the focused region is

\[ f(F|x) = \frac{f(x|F)p_1}{f(x|F)p_1 + f(x|DF)p_2}, \]
\[ = 1 - f(DF|x). \]

A pixel is labeled to F class based on the contrast value by using the following hypothesis testing constrained by a tuning parameter \( \alpha \)

\[ f(F|x) > \alpha \cdot f(DF|x). \]

The whole process can be called as pixel classification through prior learning. Here, \( \alpha \geq 1 \) is an appropriate weight chosen experimentally to reduce misclassification of pixels from DF-region to F-region, albeit it increases the overall error by misclassifying pixels belonging to F-region into DF-region. This can be easily seen from the illustration given in Fig. 3. However, it increases the retrieval accuracy which depends only on the features extracted from the segmented F-region. The requirement of \( \alpha \geq 1 \) stems from the fact that exponentially distributed clusters are poorly separable due to the maximal overlap in their distribution functions, as illustrated in Fig. 3(a). This property is studied next. Using Eq. (1), inequality 3 can be rewritten as

\[ \frac{\pi_1}{\mu_1} e^{-x/\mu_1} - \alpha \frac{\pi_2}{\mu_2} e^{-x/\mu_2} > 0. \]

or, \( x > T \) where \( T \) is the threshold shown in Fig. 3, where

\[ T = \frac{\mu_1 \mu_2}{(\mu_1 - \mu_2)} \ln \left\{ \frac{\pi_2}{\pi_1} \left( \frac{\mu_1}{\mu_2} \right) \right\}. \]

The probability that a pixel belonging to the background is wrongly classified as ROI is given by

\[ p_{fc1} = \int_T^\infty f(x|DF) \, dx, \]
\[ = \exp \left\{ \frac{-\mu_1}{(\mu_1 - \mu_2)} \ln \left\{ \alpha \left( \frac{\pi_2}{\pi_1} \frac{\mu_1}{\mu_2} \right) \right\} \right\}. \]
Thus we observe that as we increase \( \alpha \), \( pfc_1 \) decreases, signifying that the focused region contains fewer misclassified pixels. Unfortunately, we cannot keep on increasing \( \alpha \) as the number of missing pixels in the focused region (i.e., misclassified as the background class) also increase according to the following equation.

\[
p_{fc2} = \int_0^T f(x|F) \, dx,
\]

\[
= 1 - e^{-T/\mu_1},
\]

\[
= 1 - \exp \left[ \frac{-\mu_2}{(\mu_1 - \mu_2)} \ln \left( \frac{\mu_2}{\mu_1 \mu_2} \right) \right].
\]

Thus, there will be many holes in the segmented region. We found that a value of \( \alpha \) in the vicinity of 2.0 provides a good compromise. The choice of \( \alpha = 2 \) is essentially due to complete overlap of two monotone functions as distributions. Since the holes cannot be eliminated from the segmented region due to the presence of homogeneous brightness regions (these will be classified as background as the contrast is nearly zero), one may employ some post-processing technique like morphological closing for filling up some of the holes.

4. Feature extraction for CBIR

We now illustrate how the above segmentation process can be used in CBIR applications. The color feature is one of the most straightforward features utilized by humans for visual recognition and discrimination. In CBIR literature it has been widely explored by many researchers. It is relatively robust to background complication and independent of image size and orientation. We use the color histogram computed from the segmented ROI (focused) region in this study for demonstration purposes. Any other feature, either in lieu of color histogram or in conjunction with it, can also be used. The textural features can be used if we fill up the holes in the ROI through the post-processing. Thus, although all the known problems of color based CBIR continues in this study, such as sensitivity to illumination and viewing angle, and insensitivity to spatial relation between pixels, we do not pursue the choice of feature in this paper as the focus of the paper is on segmentation of low DOF images. Since each pixel in the image is described by three components in the RGB color space,
we capture the color characteristics of the image by computing three 1-D normalized histograms for R, G, and B. The computation is restricted to estimate the focused region only. Then these normalised histograms are used to compute the similarity (Euclidean distance) between the query and the database image. For experimentation purposes we use the texture retrieval method proposed in [15]. It may be noted that the computation of feature is restricted to the segmented ROI only.

5. Experimental results

Our experimental results are based on the implementation using the ground truth database of 4986 low DOF images. We created this database by collecting a good number of images from a cricket DVD video which is available online at the URL http://www.abcofcricket.com/pp/dvd-video/. Similarly we downloaded many low DOF natural and sports images from an art gallery at the URL http://www.pbase.com/galleries. The database consisted of both natural and sports images having defocus blurring. Nearly half of these images relate to various small area and wide area games like volleyball, basketball, cricket, hockey etc. Since no ground truth is available, it was constructed by other researchers in the laboratory, and is thus quite subjective from the point of view of precision and recall rate calculation. Figs. 1(a–d) shows the low DOF images from a cricket game, of a flower scene, from a hockey game, American football game, respectively. The contrast distributions of the manually cropped focused and defocused regions are shown in Figs. 2(a–d) and the overall contrast distributions of these images can be seen in Figs. 4(a–d). The contrast distribution is modeled as a mixture of two exponential distributions. We then use the EM algorithm for estimating the parameters of the mixture distribution. In all the experimental results shown ranking goes from left to right and top to bottom.

5.1. Parameter estimation

It was mentioned that the estimation of parameters is a difficult task. In order to avoid local extrema, one may have to start with good initialization and an appropriate stopping criterion. In our approach we randomly initialize...

Fig. 5. (a–d) Estimated distributions for focused (ROI) and defocused (BG) regions of the images shown in Figs. 1(a–d), and recovered from Figs. 4(a–d), respectively. The local contrast is plotted along the horizontal axis.
the values of the parameters and we are able to converge in about 1000 iterations. If the image has sizeable proportions of both focused and defocused regions, we do not face any difficulty in convergence of the EM algorithm. Fig. 5(a) shows the estimated conditional distributions of the local contrast in semilog scale for the focused region (solid line) and the defocused region (dashed line) for the image shown in Fig. 1(a). Fig. 4(a) shows the actual contrast distribution for the entire image. Similar distributions for the flower scene, an image from a hockey game and American football game are shown in Figs. 5(b–d), respectively. Figs. 6(a–d) shows the convergence of the EM algorithm for the estimation of the parameters of the mixture distribution computed for these images. In all the cases, the plots look very good and convergence does not appear to be problem. One could, in principle, stop the iterations after 300 steps.

5.2. Segmentation results

Segmentation of the image into focused and background regions is very difficult because the two distributions tend to overlap significantly over the contrast range (see Figs. 2(a–d)). This causes serious problem during the segmentation as explained in Section 3.3. The depth variation in the

![Fig. 6. Convergence of parameters for (a) the cricket, (b) the flower, (c) the hockey, and (d) the American football images. The horizontal axis denotes the number of iterations. The dot–dash curves describe the mean contrasts \( \mu_1 \) and \( \mu_2 \). Similarly dash curves represent the mixing proportions \( \pi_1 \) and \( \pi_2 \).](image)

![Fig. 7. (a and b) Segmented focused region for the cricket scene for \( \alpha = 2.0 \) and 1.0, respectively. (c and d) Results of segmentation after morphological closing for the images shown in (a) and (b), respectively. (e) Segmentation using the Blobworld approach. Here blue colored blob corresponds to the focused region.](image)
objects of interest is assumed to be small compared to the distance so that partial defocusing of ROI does not happen. We use the technique proposed in Section 3.3 to label the pixels into one of the two classes. We show results of segmentation for various values of tuning parameter \( x \). Figs. 7(a and b) show the estimated ROI of cricket image (see Fig. 1(a)) for \( x = 2.0 \) and 1.0, respectively. When \( x \) is quite small, we observe that (see Fig. 7(b)) the probability of misclassification \( p_{FC1} \) is more and many pixels of the DF-region are classified as the entities of \( F \)-region. Also at the same time few focused pixels are misclassified into DF-region. As the value of \( x \) is increased more and more focused pixels are labeled as DF-region (see Fig. 7(a)) and hence we miss quite a few focused pixels. Therefore one has to play around with the value of tuning parameter \( x \) to obtain a satisfactory segmentation result. For comparison purposes, we show the segmentation result using the Blobworld approach (see Fig. 7(e)). Here, the shaded region represents the segmented focused region. We observe that the ROI is oversegmented and many defocused pixels being misclassified as the focused. Figs. 8–10(a and b) show the segmented ROI regions and Figs. 8–10(c and d) result after morphological post-processing for low DOF scenes for flower image, a scene from a hockey and American football games, respectively. These are the results using the proposed approach. Figs. 8(e), 9(e), and 10(e) shows the corresponding segmented ROI regions obtained using the Blobworld approach. A comparison of these results clearly illustrate that the proposed technique yields much better segmentation results in low DOF images compared to the Blobworld.

5.3. Retrieval results

Qualitative evaluation of our method is carried out by visually examining the images of retrieval results. However, this can only be based on a subjective perceptual similarity since there exists no “correct” ordering that is agreed upon by all people. In all the experiments one must note that query image corresponds to the estimated ROI of the low DOF image. Fig. 11, displays the retrieved images when entire image of the cricket scene is used as the query disregarding the fact that a significant part of the scene is highly defocused. This corresponds to setting \( x = 0.0 \). We notice that very few retrieved images are relevant. Here, one can observe that the features of the blurred background in the irrelevant retrieved images such as bike rider, butterfly, bird, bear, etc., match with the features of the query

![Fig. 8](image8.png)

Fig. 8. (a and b) Segmented focused region of the flower image (see Fig. 1(b)) for \( x = 2.0 \) and 1.0, respectively. (c and d) Results of segmentation after morphological closing for the images shown in (a) and (b), respectively. (e) Segmentation using the Blobworld approach.

![Fig. 9](image9.png)

Fig. 9. (a and b) Segmented focused region of the image from a hockey game (see Fig. 1(c)) for \( x = 2.0 \) and 1.0, respectively. (c and d) Results of segmentation after morphological closing for the images shown in (a) and (b), respectively. (e) Segmentation using the Blobworld approach.

![Fig. 10](image10.png)

Fig. 10. (a and b) Segmented focused region of the American football image for \( x = 2.0 \) and 1.0, respectively. (c and d) Results of segmentation after morphological closing for the images shown in (a) and (b), respectively. (e) Segmentation using the Blobworld approach.
image (ROI) and vice versa which is, indeed undesired. This pulls down the retrieval accuracy in CBIR and calls for a ROI based technique for retrieval. We then consider the detected ROI for $\alpha = 1.0$ as the query. Retrieval results of this are shown in Fig. 12. One may notice that the number of relevant retrieved images has increased significantly and this increases the retrieval efficiency. Even here one can find a few irrelevant images. Finally we conduct experiments using query image for which the ROI is estimated with $\alpha = 2.0$. This reduces the number of background pix-

---

**Fig. 11.** Retrieval results using the entire image as the query ($\alpha = 0.0$).

**Fig. 12.** Top 20 retrieved images using features derived from ROI alone. The ROI is obtained with $\alpha = 1.0$.

**Fig. 13.** Improved retrieval performance when ROI is obtained with $\alpha = 2.0$. 

---
els being misclassified into the ROI region. Hence the estimated feature set represents the images better and our method performs extremely well. Retrieved images of this experiment can be seen in Fig. 13. We compare the performance of the proposed method with that of Blobworld, one of the most common region-based CBIR scheme. The corresponding retrieval results (using the same feature set) are shown in Fig. 14. Quite clearly, the Blobworld technique fail to match the performance of the proposed technique. This is not surprising as the Blobworld is not meant for low DOF images and it fails to perform a good segmentation of the ROI.

Similarly, experimental results for the low DOF flower scene are shown in Fig. 15 using the query detected for \( \alpha = 0.0 \). Here, again we retrieved many irrelevant images such as bird, leaves, batsman, red colored flower, etc. An improvement in retrieval accuracy is achieved using the queries estimated for \( \alpha > 1 \), results of which are shown in Figs. 16 and 17, respectively for \( \alpha = 1.0 \) and \( \alpha = 2.0 \). The results of Blobworld based retrieval are shown in Fig. 18. In another experiment we consider the query corresponding to the ROI of the image from a hockey game. Figs. 19 and 20 show the retrieved images using the queries, respectively for \( \alpha = 0.0 \) (the entire image as the query) and \( \alpha = 2.0 \). The retrieval results for the Blobworld are shown in Fig. 21. We then consider the detected ROI for the American football game as the query image. Retrieved similar images using the query for \( \alpha = 0.0 \) are shown in Fig. 22. In this case both the pixels from ROI and defocused regions participate during the feature computation. Hence, we retrieved many irrelevant images whose focused regions differ from that of the query image in terms of the color feature. These include the images occupying in 4, 5, 6, 8, 9, and 18th positions shown in Fig. 22. In addition to this we retrieved a few irrelevant images from other categories. We notice a significant improvement in the retrieval accuracy when a query computing to \( \alpha = 2.0 \) is used. Retrieved similar images are shown in Fig. 23. Here, we retrieved quite a good number of relevant images as expected. Compare this to the performance of the Blobworld shown in Fig. 24. It was earlier claimed that the ROI may have multiple objects, be them connected or disjoint, without affecting the performance of the proposed technique. In order to demonstrate this, we use another image as query where there are two footballers in focus. The corresponding retrieved images are shown in Fig. 25. We do retrieve a
large number of relevant images. This substantiates our claim that a proper CBIR scheme can, indeed, be developed by extracting the feature from the focused region only.

5.4. Performance analysis

The quantitative evaluation of the proposed scheme is measured using standard evaluation benchmarks such as
precision and recall rates. For a given query, let $T$ be the total number of retrieved images, $R_r$ be the number of retrieved relevant images, and $T_r$ be the total number relevant images. Precision $P$ is defined as $\frac{R_r}{T}$ and recall $R$ as $\frac{R_r}{T_r}$. We observe monotonic decrease in the precision rates and increase in recall rates with the increasing top $k$ number of retrieved images. Fig. 26 shows the average precision-recall curves for varying $x$. For the top 30 retrievals we obtain an average precision rate of 77.33% and an average recall rate of 93.66% when tuning parameter $x = 2.0$. When we con-
sider $\alpha = 1.0$, the precision and recall rates are reduced considerably and dropped to 56.67% and 82%, respectively. After setting the tuning parameter $\alpha = 0.0$ (i.e., when the entire image is used to compute the feature set) the precision and recall rates are further dropped down to 45.67% and 75%, respectively. The curve for $\alpha = 2.0$ indicates a gain in the retrieval accuracy when compared to the accuracies obtained for $\alpha = 1.0$ and $\alpha = 0.0$. We also evaluated the retrieval accuracy using Blobworld approach in the same plot for comparison purposes (see Fig. 26). Although
the retrieval accuracy of Blobworld approach is much better than the entire image based query, it significantly falls short of achieving the same accuracy as the proposed method does.

6. Conclusion

In this paper we proposed a framework for segmenting the focused region in a low DOF image and highlighted its use while doing the CBIR. The need for using the contrast distribution and its characteristics with reference to low DOF images have been illustrated. We presented a mixture model for the contrast distribution and discussed how one can estimate the parameters using the EM algorithm. We demonstrated that even for a difficult mixture of exponential distributions, one can obtain reasonably accurate results.

The significance of the tuning parameter $\alpha$ and its choice during pixel classification have also been discussed. We subsequently explained the use of the segmented region for image retrieval. Experimental results showed a significant improvement in the retrieval accuracy. We demonstrated the efficacy of our method by providing extensive experimental results and comparing the performance with that of the Blobworld approach. Finally, the performance of the CBIR system has been evaluated using standard evaluation benchmarks such as precision and recall rates.

As a part of future work currently we are investigating the possibility of extending the method in dealing with the objects at multiple layers of depth. Further, we plan to relate the proposed work to the growing volume of work on statistical modeling and machine learning for the purpose of image annotation. This would provide a way of annotating low DOF images.

References


Fig. 25. Retrieval results for another ROI query for American football game. Here $\alpha = 2.0$. Here two objects in the query image are in focus.

Fig. 26. Precision – recall diagrams for varying values of the tuning parameter $\alpha$ and its comparison with the Blobworld approach.


