Multi region based image retrieval system

P MANIPOONCHELVI* and K MUNEESWARAN

Department of Computer Science and Engineering, Mepco Schlenk Engineering College, Sivakasi 626 005, India
e-mail: manipoonchelvi@mepcoeng.ac.in; kmuni@mepcoeng.ac.in

Abstract.   Multimedia information retrieval systems continue to be an active research area in the world of huge and voluminous data. The paramount challenge is to translate or convert a visual query from a human and find similar images or videos in large digital collection. In this paper, a technique of region based image retrieval, a branch of Content Based Image Retrieval, is proposed. The proposed model does not need prior knowledge or full semantic understanding of image content. It identifies significant regions in an image based on feature-based attention model which mimic viewer's attention. The Curvelet Transform in combination with colour descriptors are used to represent each significant region in an image. Experimental results are analysed and compared with the state-of-the-art Region Based Image Retrieval Technique.

1. Introduction

Multimedia Information Retrieval (MIR), a solution to digital information search, is one of the challenges of current years. Web search engines, online digital data such as images/video sharing and distribution, digital libraries, image/video analysis have created a need for effective and efficient techniques to browse, search and summarize multimedia data. Content Based Image Retrieval (CBIR) is a branch of multimedia information retrieval system which deals with images. CBIR also referred as Query By Image Content (QBIC) is the application of automatic retrieval of images from a database based on the visual content such as colour, texture or shape. CBIR exploits techniques from computer vision, machine learning, database systems, data mining, information theory, statistics and psychology.

---

*For correspondence

Many CBIR systems assume each image as an entire semantic unit and gather primitive features such as colour, texture from the entire image. Because of these global features (i.e., features that are extracted from the entire image), these systems suffer a low retrieval precision (Liu *et al* 2007). Low-level global features often fail to describe the semantic content of the image. To avoid the problems associated with the global features, and to add 'semantic knowledge' to the retrieval systems, researchers have been actively working on region based approaches, techniques of fusing multiple features, probabilistically inferring the context and the techniques of providing relevant feedback to the system (Carson *et al* 2002; Wang *et al* 2001; Djordjevic & Izquierdo 2007).

A typical Region Based Image Retrieval (RBIR) system includes segmentation, feature extraction and similarity matching modules. The segmentation module divides the image into regions. Segmentation methods can be broadly classified into pixel-based, boundary-based and region-based methods. These methods include clustering, edge detection and region-growing techniques, respectively. Once the regions are obtained, features are extracted from each region. The feature extraction module gathers either low level features such as colour, texture and shape or high level semantic features which are derived from the low-level features and prior knowledge. Colour information is directly obtained from the image. The texture features are derived from statistical data such as edge, spectral coefficients. The spectral coefficients are calculated using Discrete Cosine Transform (DCT), Gabor filters, Fourier Transform (FT) and Wavelet transform (WT) (Muneeswaran *et al* 2006). The shape details are derived from Fourier descriptor, bounding box, template matching techniques or using elastic models. The extracted features are represented using appropriate descriptors and they are used while matching regions of different images.

In SIMPLIcity (Wang *et al* 2001), wavelet based approaches are used to categorize the regions into textured vs. non-texture and graphs vs. photographs. In HIRBIR (Sun & Ozawa 2005), wavelet transform is applied on image to segment distinctive regions. The feature vector (the wavelet coefficients combined with colour information) extracted from the regions is hierarchically represented for retrieval. The drawback with wavelet based approach is that it does not capture the edge information of an image effectively. A curvelet-based image retrieval approach is described in Sumana *et al* (2008). In their approach curvelet descriptors are calculated from the entire image. An approach to extract rotation invariant curvelet features is presented in Islam *et al* (2009a). The approach is better suited for textural images. Islam *et al*. (2009b) suggested a scheme to apply curvelet transform to an arbitrarily shaped object. They considered only the curvelet coefficients calculated from gray scale image. Wang *et al* (2002) generated a code book from training images to segment images into regions. The content and context of regions are represented based on colour, texture and shape. Their approach is more suitable for category based retrieval.

The proposed method is a part of the ongoing work (Manipoonchelvi & Muneeswaran 2011) of the authors. The objective of the proposed work is to identify significant regions from the colour images, represent the regions using multiple features and matching regions from query and target images to retrieve relevant images from image database. In this paper, the potential attentive regions are ranked based on the significance of the regions, thereby multiple significant regions can be considered for image retrieval.

Section 2 details how the significant regions are identified based on low level features and location cues. Section 3 details the technique of representing the significant regions. Section 4 discusses on how two images are compared with each other based on the significant regions. In section 5 of performance evaluation, we demonstrate that our proposed system performs better than the state-of-the-art system.

2. Significant region extraction

2.1 *Significant region*

In biological vision system, brain selects a subset of the visual information for high level cognitive processing by visual attention (i.e., that act of bringing salient details into focus and filter out background clutter). Experimental evidences suggest that visual attention can be guided by the processing of a limited set of basic features such as colour, orientation, size, motion and curvature (Frintrop *et al* 2010; Wolfe *et al* 2004; Hoffman & Singh 1997). Based on these findings, the proposed system identifies visually important regions in an image on which higher level vision tasks, such as object recognition, can focus.

According to Wang *et al* (2002) colour–texture classification technique is used to generate a code book from different categories of training images. The code book then is used to segment images into regions. Self and relative salient features are computed from the shape parameters extracted from each region. The computed total saliency describes perceptual importance of each region. The accuracy of their region segmentation method depends on the representative images used during training phase.

In the proposed model, firstly, we process the image to compute the saliency map and region homogeneity. Secondly, we use saliency map, location, size and homogeneity cues to locate significant regions. These cues can be extracted in parallel manner. One important advantage of our method is that we spatially group homogeneous regions to extract meaningful objects. Prior knowledge about the content of the image is not considered. Hence without the full semantic understanding of image content, the model integrates saliency map, location, and size and homogeneity values to extract meaningful significant regions.

Significant region is defined as the meaningful physical entity that is more likely to attract viewers' attention in an image. In a single image there could be more than one significant region. In our work, each significant region could be modelled by saliency value based on colour, orientation, location and homogeneity. To be specific, a significant region is defined as

$$SR = \{SV, PV, HM\},\tag{1}$$

where *SR* means significant region, *SV*, *PV* and *HM* are the salient value, proximity value and homogeneity measure of *SR*, respectively. The values of *SV* and *PV* lie within the interval [0, 1]. The magnitude of *SV* indicates a pixel's prominence with respect to surrounding pixels in a visual scene. *PV* is a measure of a pixel's distance to the center of the image. The homogeneity measure *HM* is defined as

$$HM = \{H_s^k | 1 \le k \le L, s \in S\},\tag{2}$$

$$S = \{(i, j) | 1 \le i \le M, 1 \le j \le N\},\tag{3}$$

where the size of the input image is $M \times N$, L is the number of unique labels of the segmented regions in the image. The regions with similar colour have the same label.

Given the saliency, proximity and homogeneity values, we compute optimal significant region SR^* which satisfies

$$SR^* = \underset{SR}{argmax}\, E\,(SR|SV, PV, HM; t),\tag{4}$$

here t denotes the threshold parameters used to compute the energy function in this model. The threshold parameter for *SV*, *PV* and *HM* are denoted as $t = (t_{sv}, t_{pv}, t_{hm})$, $E(.)$ is the energy function defined for a significant region *SR*. The equation (4) states that the optimal significant region is a region which maximizes the value of the energy function. The energy function is

calculated using *SV* and *PV* for a homogeneous region with a specific homogeneity measure *HM*. Elaborate discussion on how to compute the energy function is given in Manipoonchelvi & Muneeswaran (2011).

3. Significant region representation

Colour is the most dominant and distinguishing visual feature. It is relatively robust to background complication and independent of image size and orientation (Zhang 2007). The colour models specify different techniques to represent and manipulate colour values. The appropriate colour model is selected based on the specifications of an application. The colour models are broadly categorized into three groups namely device dependant colour model (ex. RGB), user dependent colour model (ex. HSV) and device independent colour model (ex. CIE Lab). The device dependent colour models are based on input, processing and output devices. The user dependent colour models are based on the human perception of colour. The device independent colour models are defined independently of devices and applications.

Colour features are obtained directly from the pixel intensities. The progressive colour descriptors are colour layout, dominant colour and colour moments. A colour layout captures both colour and spatial features. Histogram-based techniques capture concentration of the dominant colours. The colour moments capture an index containing the dominant colour features.

3.1 *Color descriptors*

Feng & Chua (2003) used a global histogram as a colour feature. An image is represented using appropriate colour space e.g., RGB, HSV or CIE Laboratory. When computing a colour histogram for an image, the colour axes are divided into number of bins and the corresponding bin's count is incremented when the colour of a pixel belongs to the bin. For a given image, the global histogram will remain the same even the image is rotated or flipped. One drawback with the global histogram is that perceptually different images with similar colour distribution will have same colour histogram. This problem can be subdued by splitting each image into smaller regions (Konstantinidis & Andreadis 2005).

In the proposed method, instead of equally splitting the image into smaller regions, only the significant regions are extracted based on the approach given in Manipoonchelvi & Muneeswaran (2011). The extracted regions represent the image content. According to this approach, an image contains one or more significant regions. The histograms based features of each of these significant regions are computed.

The HSV colour space is intended to be close to the perception of human vision system. It defines a colour by its hue, saturation and value and only hue component describes colour and the other two components describes various shades of the colour, e.g., black, white or gray. Hue represents the dominant colour as perceived by the user. Because of these properties HSV colour model is used in segmentation algorithms. Image segmentation can be done in only one colour feature, hue.

In the proposed system, the hue component is quantized into 16 bins, and the other components are quantized to four bins each. Given a digital image $I(M \times N)$ of dimension $M \times N$ is represented by the concatenated three sets of histogram based feature. The feature colour descriptor of the image I is

$$FV_I = \{FV_{r1}, FV_{r2}, \ldots, FV_{rn}\}, \tag{5}$$

where FV_I is the feature vector of the image I, FV_{ri} is feature vector of region i, and n is the number of regions.

$$FV_{ri} = \{HV_h, HV_s, HV_v\}, \tag{6}$$

where HV_h, HV_s and HV_v are the histogram based features extracted from hue, saturation and value components, respectively. The feature vectors of individual component is generally defined as

$$HV_c = \{f_c(1), f_c(2), \ldots, f_c(n_c)\}, \tag{7}$$

where n_c is the number of colour bins and $f_c(.)$ is the number of quantized colour of respective components. In case of hue, n_c is 16 and for saturation and value it is 4. Among the three components, hue, saturation and value, the hue shows more important value for human visual justification. Hence higher weight is given for hue component and 16 bins are taken for image representation. Each significant region has 24 colour feature descriptors.

Each image has one or more significant region. When multiple regions are distinguished in an image then the feature vector for each region are computed and stored separately. During image matching process they are separately handled.

3.2 *Curvelet descriptors*

Recent researches on multi-resolution analysis have found that curvelet captures texture properties, like, curves, lines and edges, more accurately than other multi-resolution based approaches (Sumana *et al* 2008; Aroussi *et al* 2009). In our work, we extract texture features of the identified significant regions using curvelet transform.

The curvelet transform is a multi-scale directional transform which allows an optimal sparse representation of objects with discontinuous contours. For a digital image $I(M \times N)$ of dimension $M \times N$, the curvelet coefficients are obtained by implementing the Fast Discrete Curvelet Transform (FDCT) (Candès *et al* 2006). FDCT is computed by decomposing the image into a set of directional sub-band images at multiple scales. The curvelet coefficient c of an image $I(m, n)$, $1 \le m \le M$, $1 \le n \le N$, at scale 2^s, orientation θ angles and position $\kappa = (k1, k2)$ is obtained by

$$c(s, \boldsymbol{\theta}, \boldsymbol{\kappa}) = IFFT(FFT(I[m, n])) * FFT(\boldsymbol{\varphi}_{s,\theta,\kappa}[m, n]), \tag{8}$$

$$\boldsymbol{\varphi}_{s,\theta,\kappa}(m, n) = R_\theta^T * U_s(m, n)e^{-2\pi i(k_1 m/M_{s\theta} + k_2 n/N_{s\theta})}, \tag{9}$$

where *IFFT* is Inverse Fast Fourier Transform, *FFT* is Fast Fourier Transform, φ is curvelet waveform or mother curvelet, $s(= 0, 1, \ldots)$ is scale factor, θ is orientation in the interval $0 \le \theta \le 2\pi$. The curvelet coefficients in each sub-band is symmetric in nature i.e., the curvelet at angle θ is same as that of the coefficient at angle $(\theta + \pi)$ in the frequency domain. Taking advantage of this property of curvelet, only half of the sub-bands at each scale are considered for feature calculation.

The curvelet coefficients at each scale, orientation and location are normalized by dividing each coefficient by the corresponding sub-band's area. Later, a threshold is employed on the coefficients to remove noise. Curvelet descriptors are computed as mean ($\mu_{s\theta}$) and standard deviations ($\sigma_{s\theta}$) of each selected sub-bands.

$$\mu_{s\theta} = \frac{\sum_m \sum_n |c_{s\theta}(m, n)|}{m * n}, \tag{10}$$

$$\sigma_{s\theta} = \frac{\sqrt{\sum_m \sum_n (|c_{s\theta}(m, n)| - \mu_{s\theta})^2}}{m * n}. \tag{11}$$

Curvelet descriptors are computed and indexed for every region in the images in the database to speed up the retrieval process.

4. Region matching

4.1 *Matching colour features*

Given query image is matched with every images in the database. The images which have higher similarity or lower distance with the query images are labelled as retrieved images. Histograms of images/regions are matched by many techniques. The standard techniques are bin-by-bin Euclidean distance, Earth Mover's Distance (EMD) and Histogram Intersection (Rubner *et al* 2000; Swain & Ballard 1991). In the proposed system, the colour similarity between two images is measured using histogram intersection technique.

The histogram similarity of two regions Q and T that are represented by the respective feature vectors FV_t and FV_q is defined as

$$SM(T, Q) = \frac{\sum_{i=1}^{n} \min\left(FV_t(i), FV_q(i)\right)}{\min\left(\sum_{i=1}^{n} FVt(i), \sum_{i=1}^{n} FVq(i)\right)}, \tag{12}$$

where n is the number of features in the feature vector. The numerator in equation (12) denotes the histogram intersection i.e., the number of features from the region Q that have corresponding features of the region T. The histogram intersection is normalized to the interval [0, 1] by the total number of pixels in the smallest region. The similarity measure (*SM*) increases when both the regions have same set of colour features. If both the regions are identical or either one of the region is contained in the other region the *SM* becomes 1. If both the regions are dissimilar the *SM* becomes 0. The distance between the two regions is defined as

$$D_{Colour}(T, Q) = 1 - SM(T - Q). \tag{13}$$

If both the regions have same set of colour descriptors, the *SM* is high and the distance is low.

4.2 *Matching curvelet features*

The significant regions are normalized to 64×64 pixels and 20 curvelet descriptors are calculated. The similarity of each region based on its texture property is computed using Euclidean distance.

$$D_{Curve}(T, Q) = \sqrt{\left(\sum_{i=1}^{2n} \left(T_{cd_i} - Q_{cd_i}\right)^2\right)}, \tag{14}$$

where $D_{Curve}(T, Q)$ is the distance between two regions Q and T, Q_{cd} is curvelet descriptor of query region, T_{cd} the curvelet descriptor of significant region in the target image in the database and n is the number of sub-bands. The small value of the distance indicates that both the region have the same set of features. When both the regions are same $D_{Curve}(T, Q)$ becomes 0 and it becomes 1 when both the regions are completely dissimilar.

4.3 *Fusing colour and texture features*

Both query and target images have one or more significant regions. To match regions of query and target images one-to-one, one-to-many, many-to-one or many-to-many approaches can be employed. In all these approaches the similarity measure between each region in the query image and each region in the target image is computed. The total distance between two regions is computed as

$$D(T, Q) = D_{Colour}(T, Q) + D_{Curve}(T, Q). \tag{15}$$

The distance between the query image and the target image is computed as the minimum distance among all possible query and target region pairs.

Let us assume that query image (I_q) contains M significant regions and the target image (I_t) contains N regions. The similarity between these two images is computed as

$$D\left(I_t, I_q\right) = min\left(D\left(T_i, Q_j\right)\right), \tag{16}$$

where $D\left(T_i, Q_j\right)$ is the distance between target region T_i and query region Q_j, $i = 1..M$ and $j = 1..N$, M is the number of regions in the query image and N is the number of regions in the target image. The distance between each pair of query and target image is computed and the images in the database are sorted out accordingly. The distance measure is used for filtering and ranking the resultant images.

5. Performance evaluation

5.1 *Performance metric*

The objective evaluation of the proposed approach is carried out based on precision, recall and F-measure. These three measures are computed out of ranked and unranked resultant images. Precision measures the accuracy of the retrieval. Precision is the ratio of retrieved images that are relevant to the query image.

$$P = \frac{Number\ of\ relevant\ images\ retrieved}{Total\ number\ of\ images\ retrieved}. \tag{17}$$

Recall measures the robustness of the retrieval. It is defined as the ratio of relevant images in the database that are retrieved in response to a query.

$$R = \frac{Number\ of\ relevant\ images\ retrieved}{Total\ number\ of\ relevant\ images\ in\ DB}. \tag{18}$$

The images in the database are manually categorized into different semantic groups for the evaluation of the result. The resultant image is counted as relevant image if and only the semantic group of both the query and the resultant image are the same.

F-measure is the unified performance measure.

$$F_{\alpha}Measure = \frac{(1 + \alpha)Precision \times Recall}{(\alpha \times Precision) + Recall}, \tag{19}$$

where α indicates the importance of precision over recall. We use $\alpha = 0.25$ for emphasizing precision over recall.

5.2 *Experimental results*

The experiments are conducted on Coral database of about 10,000 images with size of 384 ×
256 or 256 × 384. The images are classified into semantic categories and each category con-
tains 100 images. To compare the results with the existing algorithms, ten categories (Flower,
Mountain, Food, Buildings, Bus, Africa, Beach, Horse, Dinosaur, and Elephant) are considered.
Every image in the selected 10 categories is tested as a query. The same set of parameters is used
throughout the experiments. The average performance of each image category is computed in
terms of precision and F-Measure.

Each image is resized to 256 × 256 and decomposed at 5 levels using CurveLab 2.1.2. To
calculate Gabor features we used 4 frequencies of 6 orientations using Gabor Feature Toolbox
version 1.0.0 of Ilonen & Kämäräinen (2006).

The results published in Sun & Ozawa (2005) are taken to compare the effectiveness of the
proposed system over SIMPLIcity (Wang *et al* 2001) and HIRBIR (Sun & Ozawa 2005). As
stated in Sun & Ozawa (2005), the retrieval precision is calculated as the averaged value in the
top 30 images and the results are given in figure 1a. The empirical results show that the proposed
approach performs better than the methods presented in Wang *et al* (2001), Sun & Ozawa (2005).

The retrieval precision and F-Measure of the proposed Multi Significant Region based
approach (MSR), Single Significant Region based approach (SSR) (Manipoonchelvi &
Muneeswaran 2011), Features extracted from entire image and Curvelet Transformed based
approach (FCT) (Sumana *et al* 2008) and Multi Significant Region and Gabor Transform based
approach (MSRG) are calculated as the averaged value in the retrieved images. The performance
comparison data are given in figures 1b and 1c.

We compared the performance of our method with semantics-based image retrieval by region
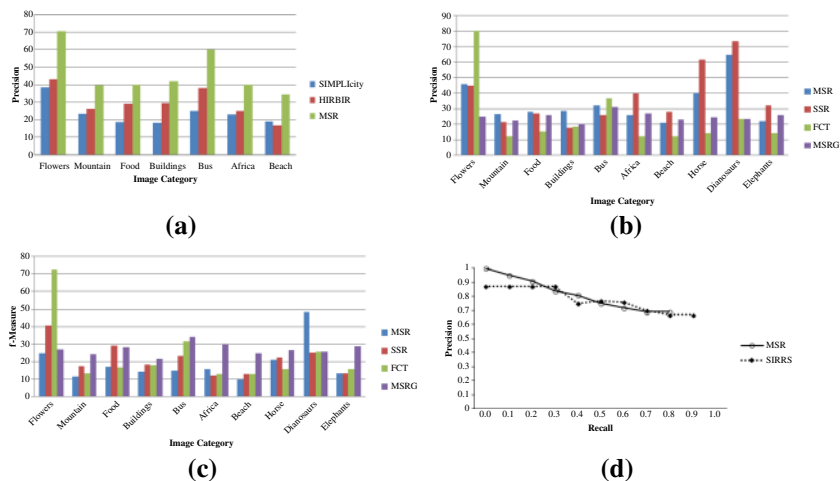saliency (SIRRS) approach proposed in Wang *et al* (2002). We used 3925 images which are

Figure 1. Performance Evaluation using Corel Database (**a**) comparison of the state-of-the-art techniques,
SIMPLicity and HIRBIR, with the proposed MSR technique (**b**) Precision based comparison of proposed
MSR and Single Significant Region based (SSR) approaches with Curvelet Transform from the entire image
(FCT) approach and Gabor Transform based approach (MSRG) (**c**) F-Measure based comparison of MSR
and SSR approaches with FCT and MSRG approaches (**d**) Precision-Recall curve based comparison of the
proposed MSR technique with SIRRS technique.

manually categorized into 9 semantic categories namely Sky, Water, Tree and Grass, Falls and River, Flower, Earth and Rock and Mountain, Ice and Snow and Mountain, Sunset and Night. We calculated and plotted the interpolated precision-recall curve in figure 1d. The Area Under Curve (AUC) for the proposed method is 0.79 and for SIRRS is 0.77. The average precision of the proposed method (81%) is higher than that of the SIRRS (69%). In other words the proposed method is more accurate than SIRRS. The SIRRS method has employed off-line learning to segment an image. Without having the expensive learning procedure, our proposed method produces higher retrieval precision than SIRRS.
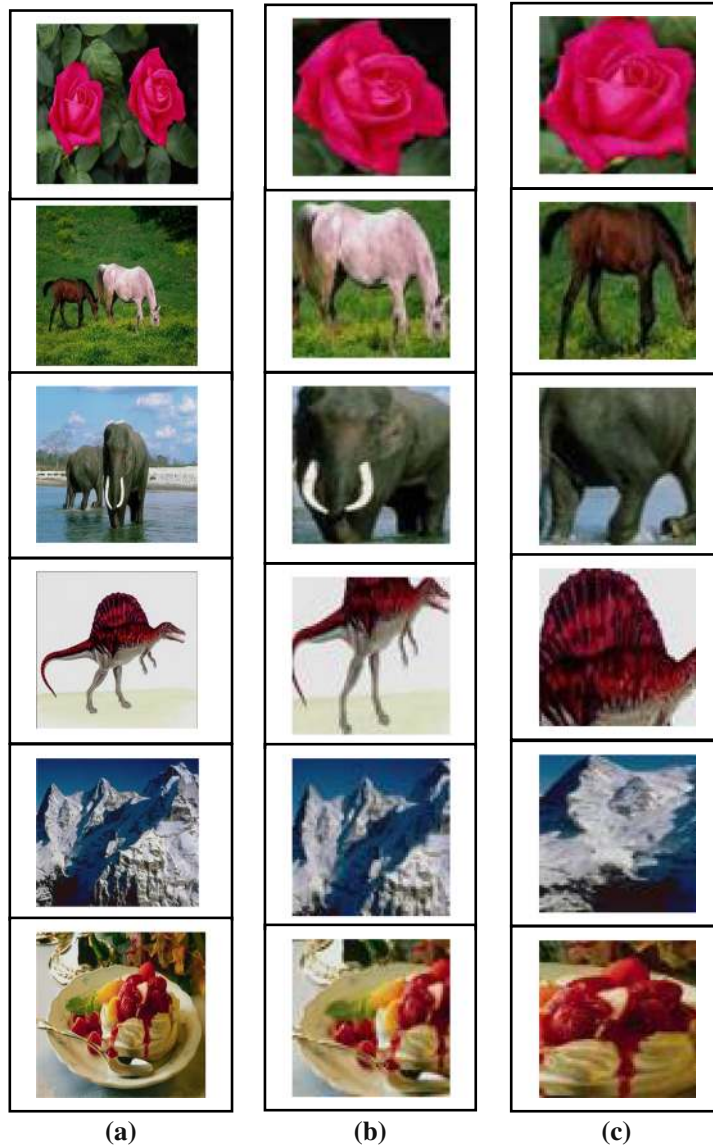

Figure 2. Significant regions: (**a**) the input image, (**b**) the primary significant region, (**c**) the secondary significant region.

The advantages of proposed methods compared to the existing systems are:

- In Sumana *et al* (2008), curvelet feature is used for image retrieval and the curvelet coefficients are extracted from the entire image. In the proposed system, the features are extracted only from the identified significant regions.
- In SIMPLIcity (Wang *et al* 2001), wavelet and k-means clustered-based image segmentation algorithm is used. The region segmentation algorithm employed in the proposed system uses semantic cues such as location, size and homogeneity. Curvelet transform-based features are used to capture texture properties.
- The salient region extraction process can be executed in parallel manner and colour and curvelet features can also be extracted in parallel manner for better performance.
- Unlike SIRRS (Wang *et al* 2002), no prior knowledge about the content of the image is required.

In the experiments, we found that significant region extraction results did not meet the users satisfaction well in few images, especially for the images with the complicated background, semantic objects composed of many smaller parts or images with very big object (occupies more than half of the image area) in the scene.

The identified significant regions are ranked based on their significance value. The first two (primary and secondary) significant regions are shown in figure 2. While regions are accurately segmented, that is any object in the image is not partially segmented then SSR lists more precise results than proposed MSR approach in this paper. That is, wherever a single semantic object is not properly extracted due to its higher textured and inhomogeneous surfaces, the performance of multi region based approach is relatively better than the single region based approach.

The identified significant regions are evaluated by judging whether the identified regions form meaningful/important objects in each image. We performed subjective assessment to check whether the identified significant regions match with human expectations. For the evaluation purpose, we selected two categories (Horse and Elephant), each containing 100 images. Each image contains one or more objects (either Horse or Elephant) of different sizes at different positions. Human observers rated the results obtained by the proposed method as perfect/partial/failed based on the accuracy of the identified significant regions. The 'Perfect' rating was given when the important objects were not missed and the full shape of the objects was captured. The 'Partial' rating was given when the extracted regions were not very accurately identified. When the significant objects were not captured by the algorithm, the result was marked as 'Failed'. The summary of the subjective evaluation is given in table 1. From the table we can observe that one category of images ('Horse') achieved better accuracy than that of the other category. The reason is the majority of the images which belong to 'Elephants' category does not have visually distinctive foreground and background. That is, the dominant objects in the images almost blend

Table 1. The subjective assessment for the proposed approach.

	Statistical results are presented in %		
	Perfect	Partial	Failed
Horse	79	19	2
Elephant	53	41	6
Average	66	30	4

with the background and hence the objects are less distinguishable from the background. In these scenarios, high level semantics can be used to segment meaningful regions in the images.

6. Conclusion

This paper proposed an approach which automatically extracts significant regions/objects in the user submitted query image and retrieves a set of colour images from the image database based on their similarity to the content of the query image. The colour and texture features are obtained from the identified multiple significant regions and represented using histogram and curvelet transform, respectively. The region identification procedure is also similar to human perception. The empirical measures suggest that the proposed method is better suited for cluttered and complex images in which the foreground is not prominent from the background images. Our future work is to consider shape adaptive representation of the significant regions.

References

Aroussi E, Ghouzali M, Hassouni S E, Rziza M and Aboutajdine M 2009 Curvelet-based feature extraction with B-LDA for face recognition. *Proc. Int. Conf. Computer Systems and Appl. (AICCSA)*, 444–448

Candès E J, Demanet L, Donoho D L and Ying L 2006 Fast discrete curvelet transforms. *SIAM J. Multiscale Model. Simul.* 5(3): 861–899

Carson C, Belongie S, Greenspan H and Malik J 2002 Blobworld: Image segmentation using expectation-maximization and its application to image querying. *IEEE Trans. Pattern Anal. Mach. Intell.* 8(8): 1026–1038

Djordjevic D and Izquierdo E 2007 An object- and user-driven system for semantic-based image annotation and retrieval. *IEEE Trans. Circuits and Syst. Video Technol.* 17(3): 313–323

Feng H and Chua T S 2003 A boostrapping approach to annotating large image collection. *Proc. Workshop Multimedia Information Retrieval in ACM Multimedia*, 55–62

Frintrop S, Rome E and Christensen H I 2010 Computational visual attention systems and their cognitive foundations: A Survey. *ACM Trans. Appl. Percept.* 7(6): 1–39

Hoffman D D and Singh M 1997 Salience of visual parts. *Cognition* 63: 29–78

Ilonen J and Kämäräinen J 2006 Simplegabor - Multiresolution Gabor Feature Toolbox, Gabor Feature Toolbox version-1.0.0 http://www2.it.lut.fi/project/simplegabor/downloads/src/simplegabortb/

Islam M M, Zhang D and Lu G 2009a Rotation invariant Curvelet feature for texture image retrieval. *Proc. IEEE Int. Conf. Multimedia and Expo*, 562–565

Islam M M, Zhang D and Lu G 2009b Region based color image retrieval using Curvelet transform. *Proc. ACCV* 2: 448–457

Konstantinidis K and Andreadis I 2005 Performance and computational burden of histogram based color image retrieval techniques. *J. Comp. Methods in Sci. and Eng* 4: 141–147

Liu Y, Zhang D, Lu G and Ma W Y 2007 A survey of content-based image retrieval with high-level semantics. *Pattern Recognition* 40: 262–282

Manipoonchelvi P and Muneeswaran K 2011 Significant region based image retrieval using curvelet transform. *Int. Conf. Recent Advancements in Electrical, Electronics, and Control Eng. (ICONRAEeCE)*, 291–294

Muneeswaran K, Ganesan L, Arumugam S and Soundar K R 2006 Texture image segmentation using combined features from spatial and spectral distribution. *Pattern Recognition Lett.* 27(7): 755–764

Rubner Y, Tomasi C and Guibas L J 2000 The Earth Mover's Distance as a Metric for Image Retrieval. *Int. J. Comput. Vision* 40(2): 99–121

Sumana I, Islam M, Zhang D S and Lu G 2008 Content Based Image Retrieval using Curvelet Transform. *Proc. IEEE Int. workshop Multimedia Signal processing*, Australia, 11–16

Sun Y and Ozawa S 2005 A hierarchical approach for region-based image retrieval. *Proc. IEEE Int. Conf. Systems, Man and Cybernetics*, 1117–1124

Swain M J and Ballard D H 1991 Color Indexing. *Int. J. Comput. Vision* 7(1): 11–32

Wang J Z, Li J and Wiederhold G 2001 SIMPLIcity: Semantics-sensitive Integrated Matching for Picture libraries. *IEEE Trans. Pattern Anal. Mach. Intell.* 23(9): 947–963

Wang W, Song Y and Zhang A 2002 Semantics-Based Image Retrieval by Region Saliency. *Image and Video Retrieval, Lecture Notes in Comput. Sci.* 2383: 29–37

Wolfe J M, Horowitz T S, Kenner N, Hyle M and Vasan N 2004 How fast can you change your mind? The speed of top-down guidance in visual search. *Vision Res.* 44: 1411–1426

Zhang Y J 2007 *Semantic-Based Visual Information Retrieval*, IRM press