Content Based Image Retrieval Using Region Labelling

J. Naveen Kumar Reddy, Chakravarthy Bhagvati, S. Bapi Raju, Arun K. Pujari, and B.L. Deekshatulu

Dept. of Computer and Information Sciences
University of Hyderabad, Hyderabad 500046

Abstract. This paper proposes a content based image retrieval system that uses semantic labels for determining image similarity. Thus, it aims to bridge the semantic gap between human perception and low-level features. Our approach works in two stages. Image segments, obtained from a subset of images in the database by an adaptive k-means clustering algorithm, are labelled manually during the training stage. The training information is used to label all the images in the database during the second stage. When a query is given, it is also segmented and each segment is labelled using the information available from the training stage. Similarity score between the query and a database image is based on the labels associated with the two images. Our results on two test databases show that region labelling helps in increasing the retrieval precision when compared to feature-based matching.

1 Introduction

Research in Content Based Image Retrieval (CBIR) steadily gained momentum in recent years consequent to the dramatic increase in the volume of digital images. Image databases containing thousands and sometimes millions of images are available in many fields such as remote sensing, medical imaging and biometrics (e.g., fingerprints). The main difference between data retrieval from a conventional database and CBIR is that the former is based on predefined keywords (or keys) associated with each stored record, while the latter utilises visual cues. A good survey of CBIR field may be found in the review paper by Sameer Antani, Rangachar Kasturi and Ramesh Jain[1].

CBIR systems mainly use colour[2,3,4,5], texture[6,7,8], shape[9,10,11] and other low-level features to assess similarity while human beings rely on high-level symbolic (e.g., chair, Taj Mahal) and abstract (e.g., war, happy occasions) concepts. Such a difference between computer and human perception of similarity, often called the *semantic gap*, led researchers into exploring methods that fall broadly into three categories: relevance feedback, modeling human perception and linguistic indexing.

Relevance feedback is an on-line learning strategy where user provides feedback on the retrieved images which adapts the response from the CBIR system. Usually, the user responses are used either to modify the query or the similarity measures[12,13]. Several attempts have been made to incorporate human

P. Kalra and S. Peleg (Eds.): ICVGIP 2006, LNCS 4338, pp. 872–881, 2006.

perception into CBIR systems by developing computational models for early human vision[14,15,16] or similarity functions consistent with human perception [17,18,19]. In lingusitic indexing, images are categorized into different types and annotated (or labelled) by using a trained classifier. The problem of CBIR is thus reduced to the problem of text-based retrieval. Examples of such work are found in the paper by Chen and Wang[20].

There are two major approaches to linguistic indexing. The first is region-based and initially segments images into regions. If the segmentation process is ideal each region is treated as a semantically meaningful object and retrieval is based on similarity between regions. UCSB NeTra[7], Berkeley's Blobworld[21] and Stanford's SIMPLIcity[22] are examples of such systems. Integrated Region Matching (IRM) [23] (and used in [22]) and its enhancement using fuzzy membership functions for more perceptual segmentation and region matching[24] illustrate how sophisticated region-based similarity measures may be defined for CBIR applications.

The second approach is to organize the digital library in a semantically meaningful manner using image classification. Such classification is also useful to index images automatically. Unfortunately, there is no effective method yet to obtain good semantic categorization from low-level features. A compromise is manual annotations, which are potentially subjective and ambiguous and sometimes difficult because image data is rich in detail. In specific domains, however, classification provides a powerful set of semantic features for CBIR.

In this paper, we describe a CBIR system based on semi-automatic region labelling for remote sensing images. Remote sensing images have the advantage that their classification is well-researched, and well-defined with several standard schemes in existence. We also used the same method on a miscellaneous database containing different categories of objects such as flowers, aeroplanes, flags, etc. Our results indicate that region-labelling, even though it is sometimes incorrect, leads to improved precision in retrieving images.

The rest of the paper is organized as follows. Section 2 describes feature extraction and segmentation. Section 3 describes the two-stage segment labelling process and subsequent query processing. Section 4 illustrates the results on two test databases and Section 5 concludes the paper.

2 Feature Extraction and Segmentation

Our system segments images using an adaptive k-means algorithm based on colour and texture features. We follow the same approach described in [22] and [23]. An image is divided into 4×4 blocks from which three colour and three texture features are computed. The colour features are the average L, U and V components where LUV colour space is used for its perceptually uniform properties. The three texture features are obtained from Daubechies-4 wavelet transform on the L component. The 4×4 block is decomposed into four 2×2 frequency bands after a one level wavelet transform. The HL texture feature f_{HL} is then given by

$$f_{HL}(k,l) = \left(\frac{1}{4} \sum_{i=0}^{1} \sum_{j=0}^{1} C^2_{k+i,j+l} \right)^{\frac{1}{2}}$$

where C_{00}, C_{01}, C_{10} and C_{11} are the four values in each 2×2 block. The HH and LH band features are similarly calculated.

k-means algorithm is used for segmenting the images by clustering the feature vectors. If the number of 4×4 blocks in the image is B, the goal of k-means algorithm is to partition the set of feature vectors $f_i, 1 \leq i \leq B$ into k groups with means $\overline{f}_1, \overline{f}_2, \ldots, \overline{f}_k$ such that

$$D(k) = \sum_{i=1}^{L} \min_{1 \leq j \leq k} (f_i - \overline{f}_j)^2$$

is minimized.

We adaptively choose the number of clusters by starting with $k = 2$ and increasing it until one of the following three criteria is met: $D(k)$ is below a specified threshold indicating that the clusters are tight; change in $D(k)$ between consecutive iterations is below a threshold indicating convergence; and, k exceeds an upper bound indicating too large a number of clusters.

3 Region Labelling

Region labelling step assigns a semantic label to each segment obtained from k-means clustering. The labelling is done in two stages. A subset of images from the database that illustrate the semantic concepts important in querying is chosen in the first, manual or training, stage. The different segments in the images are shown to the trainer and are assigned labels manually. As there exist many regions that have identical labels, the average feature vector of such regions along with the associated label are stored in a label database. Some sample regions and labels are shown in Figure 1 from a test database of approximately 1200 low-resolution remote sensing images obtained from the National Remote Sensing Agency (NRSA), Hyderabad. An example of a segmented image that is labelled manually is shown in Figure 2.

In the second stage, all the segmented images in the database are automatically assigned labels utilizing the label database created in the training stage. The feature vector corresponding to an unlabelled segment is compared with the mean feature vectors in the label database. The label associated with the nearest feature vector in the label database is assigned to the unlabelled region. An example of automatic labelling is shown in Figure 3. It may be noticed that the labelling is not fully correct in that snowy regions are labelled as a *water body*. Such inaccuracies have an impact on the precision and recall values of the CBIR system as we shall see in the next section.

When a query image is given, it is segmented and automatically labelled using the approach described above. Image similarity is measured on the labels and not

Color Name	Color	Interpretation
Dark Red		Dense Vegetation
Red		Normal Vegetation
Light Red		Sparse Vegetation
Pure Black		Pure Water Body
Black		Water Body with less sediments
Gray		Water Body with less sediments
Dark Blue		Water Body with some sediments
Blue		Water Body with some sediments
Grayish Blue		Water body with more sediments
Cyan		Water Body with more sediments

Fig. 1. Image segments and associated labels

on the feature vectors leading to retrieval using semantics rather than low-level characteristics. Similarity score is given by

$$S = \frac{n_m}{n}$$

where n is the number of regions in the query image and n_m is the number of identical labels between the query and the database images. S ranges from 0 to 1.0. Many variants of the above similarity score exist in literature[25] although they have not been tried yet by us.

4 Experimental Results

We initially experimented with a database of remote sensing images from NRSA, Hyderabad. The retrieved results for a query image showing waterbody along with some clouds are shown in Figure 4. Similarity score is based on the labels automatically assigned to the images (as explained in the previous section). It may be seen that all the top eight images are relevant. However, several of the images contain clouds while there are many other images in the database that are not cloudy.

More sophisticated queries allowed by our system specify the *location, extent* or *absence* of a class in conjunction with a query image. In this case, additional

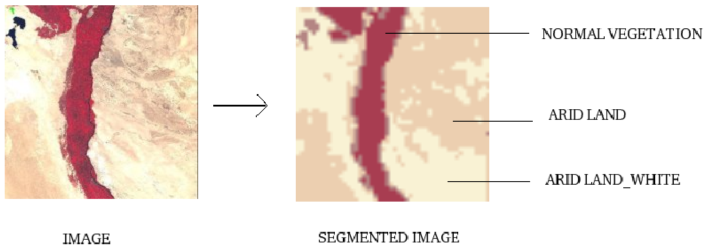

Fig. 2. Example of a segmented image that is manually labelled

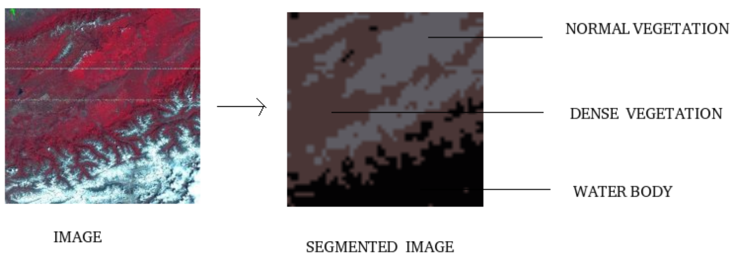

Fig. 3. Example of a segmented image that is automatically labelled

information about the bounding box and area of the segment is included along with the similarity measure. One example of such queries is extremely useful to NRSA and involves retrieving *cloud-free* images similar to a given query. Figure 5 shows the results for the same query image of Figure 4 but with the additional condition that the results do not contain cloudy regions. It may now be noticed that the query image is not the first image to be retrieved but the third. The first two images are less cloudy leading to higher similarity score than for the query image itself because its cloudy regions lower the similarity with the query. Overall, the resulting images are of better quality than the earlier set but some of them still contain clouds. The reason is that the wispy clouds seen are misclassified as water bodies with sedimentation and are hence retrieved.

We also experimented with a much larger and more varied database containing approximately 8000 images. These show objects such as flowers, aeroplanes, sunsets, fireworks and others. Figure 6 shows an example query image of fireworks and the retrieved images. Again, all the top eight images are relevant. We compared our results against the Integrated Region Matching (IRM) method proposed in [23]. The results from the IRM method are shown in Figure 7. It may be seen that three of the images do not show fireworks and are therefore irrelevant.

A more detailed comparison of the performance of the region-labelling method and IRM is shown in Figure 8. We took 10 query images each from seven categories — remote sensing (C1), flags (C2), flowers (C3), landscapes (C4),

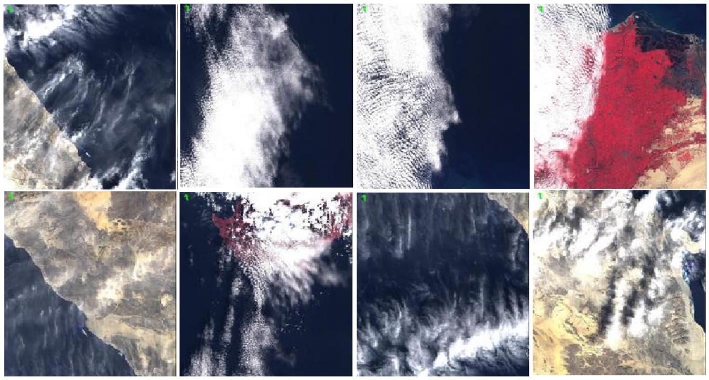

Fig. 4. Retrieved images for *waterbody* class using region-labelling method. Query is the first image.

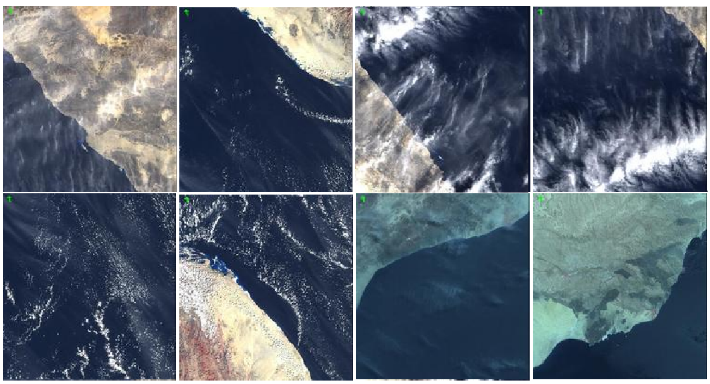

Fig. 5. Retrieved images when an additional *cloud-free* constraint is specified. Query is the same as in Figure 4.

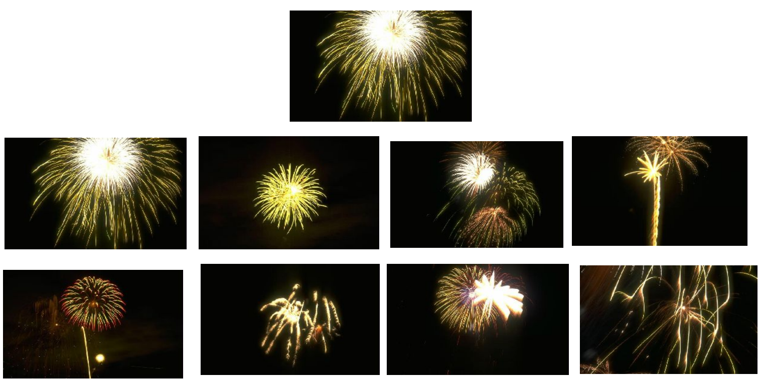

Fig. 6. Query results using region-labelling for the fireworks image shown at the top

Fig. 7. Query results for the same query image as in Figure 6 using IRM method

automobiles (C5), aeroplanes (C6), and fruits and vegetables (C7) — and computed the average precision for each category. We first used region-labelling and then the IRM method. From the Figure 8 it may be seen that for Category *C1*, the average precision for region-labelling method (identified as *labelling* in the figure) is approximately 0.9 while it is about 0.8 for IRM. The average precision for all categories (except *C4*, i.e., landscape images) is higher than 0.7 indicating good overall performance.

It may also be seen that for all categories except *C3*, i.e. flower images, region-labelling results in a higher average precision than IRM.

Finally, we show a recall-precision graph (Figure 9) that compares the performance of the region-labelling approach with IRM and its fuzzy extension UFM in retrieving remote sensing images. It may be seen that the region-labelling approach is better than both IRM and UFM.

An image contains many colours and regions, but a human user generally focusses on only a few high level concepts while ignoring the finer details. Consequently, in our system, we restricted the maximum number of segments in an image to six. The performance of our system on low-resolution remote sensing images and on images containing flags, fireworks and other such easily distinguishable objects justifies the choice of a small value for the number of segments. We noticed that our system gives consistently high precision and recall when the images contain one or two dominant objects.

We used simple colour and texture features in our approach. They may not be adequate to distinguish between certain semantic classes. An example is Figure 5 where images containing clouds are present in the results even though the query specified otherwise. The 6 features are not capable of distinguishing between water containing large levels of sedimentation and clouds, both of which are greyish white in colour and contain no texture.

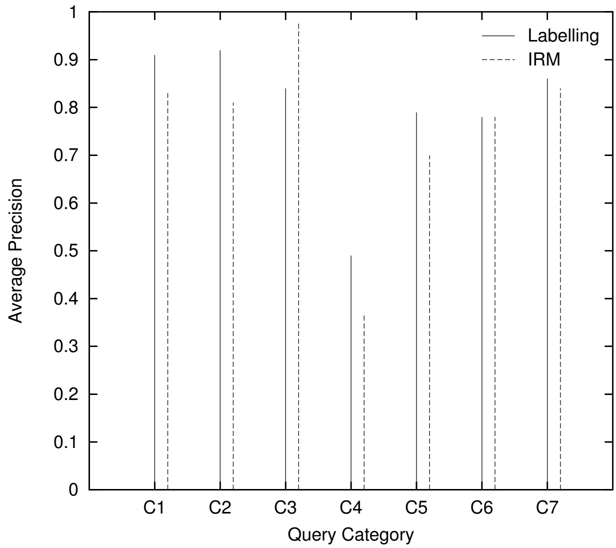

Fig. 8. Comparison of average precision between region labelling and IRM methods over 7 categories of images

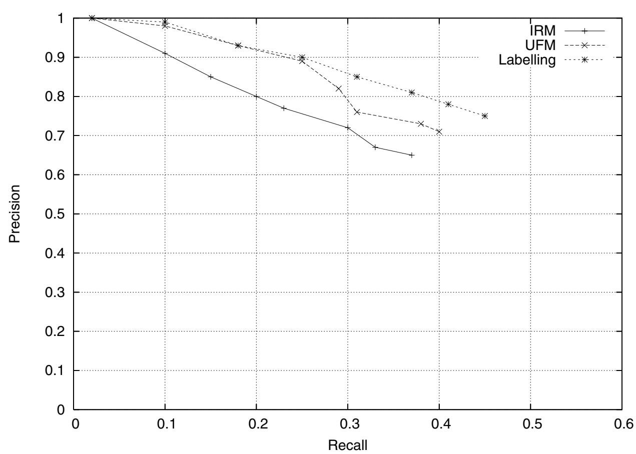

Fig. 9. Recall-precision graph comparing region labelling with IRM and UFM

Sometimes, however, it is not easy to label the segments in an image. For example, *landscapes* are particularly difficult if they involve a panoramic view of the countryside with few dominant features. There are two problems with such images: there are more than 6 semantically meaningful segments, and it is not easy to associate a label with low-level features. The latter problem causes our classification method based on nearest neighbour approach to fail.

To summarise, our proposed CBIR system based on supervised classification gives higher precision in retrieval when images contain a small number of meaningful segments and a few dominant objects. In other cases, either the segmentation or the labelling becomes inaccurate. Inaccuracy in segmentation does not seriously affect the performance as the similarity measure is based mainly on labelling and not on region properties. However, poor segmentation can lead to inaccurate labelling. Labelling has a direct impact on precision and good classification algorithms should be used.

5 Conclusion

In the paper, we presented a CBIR system that retrieves images based on high-level semantics that are assigned using a trained classifier. The results indicate that even our simple approach based on only 6 colour and texture features in conjunction with a supervised nearest neighbour classification scheme can lead to higher precision in retrieving several types of images. Our approach, demonstrated on remote sensing images, may be extended to querying other image databases provided that the images contain a small number of regions and distinct features that act as cues for matching.

References

1. Antani, S., Kasturi, R., Jain, R.: A Survey on the Use of Pattern Recognition Methods for Abstraction, Indexing and Retrieval of Images and Video. Pattern Recognition **35** (2002) 945–965
2. Fuertes, J.M., Lucena, M., Blanca, N.P.d.l., Chamorro-Martinez, J.: A Scheme of Colour Image Retrieval from Databases. Pattern Recognition Letters **22** (2001) 323–337
3. Kankanhalli, M.S., Mehtre, B.M., Huang, H.Y.: Colour and Spatial Feature for Content-Based Image Retrieval. Pattern Recognition Letters **20** (1999) 109–118
4. Pass, G., Zabih, R., Miller, J.: Comparing Images Using Color Coherence Vectors. In: Proc. of the ACM Multimedia'96 Conference. (1996) 65–73
5. Prasad, B.G., Gupta, S.K., Biswas, K.K.: Region-Based Image Retrieval Using Integrated Color Shape and Location Index. Computer Vision and Image Processing **94** (2004) 193–233
6. Li, M., Chen, Z., Zhang, H.J.: Statistical Correlation Analysis in Image Retrieval. Pattern Recognition **35** (2002) 2687–2693
7. Ma, W.Y., Manjunath, B.S.: NETRA: A toolbox for navigating large image databases. In: Proc. IEEE Int. Conf. on Image Processing, Santa Barbara, California (1997) 568–571
8. Unser, M.: Texture Classification and Segmentation Using Wavelet Frames. IEEE Trans. on Image Processing **4** (1995)
9. Zhang, J., Zhang, X., Krim, H., Walker, G.G.: Object Representation and Recognition in Shape Spaces. Pattern Recognition **36** (2003) 1143–1154
10. Badawy, O.E., Kamel, M.: Shape based image retrieval applied to Trademark images. Int. J. of Graphics and Image Processing **2** (2002) 1–19

11. Jain, A.K., Vailaya, A.: Shape-Based Retrieval: A Case Study with Trademark Image Databases. Pattern Recognition **31** (1998) 1369–1390
12. I.J.Cox, M.L.Miller, T.P.Minka, T.V.Papathomas, Yianilos: The bayesian image retrieval system, PicHunter: theory,implementation and psychophysical experiments. IEEE Trans. Image Processing **9** (2000) 20–37
13. Kherfi, M.L., Ziou, D., Bernardi, A.: Learning from Negative Example in Relevance Feedback. IEEE Trans. on Image Processing (2002)
14. Lu, G., Phillips, J.: Using perceptually weighted histograms for colour-based image retrieval. In: Proc. Fourth Int. Conf. on Signal Processing. (1998)
15. Mojsilovic, A., Kovacevic, J., Hu, J., Safranek, R.J., Ganapathy, S.K.: Matching and retrieval based on the vocabulary and grammar of color patterns. IEEE Trans. on Image Processing **9** (2000)
16. Nathalie, G., Heve, L.B., Jeanny, H., Anne, G.D.: Towards the introduction of human perception in natural scene classification system. IEEE Trans. on Neural Networks and Signal Processing (2002)
17. Rubner, Y., Tomasi, C.: Perceptual Metrics for Image Database Navigation. Kluwer Academic Publishers (2001)
18. Celebi, M.E., Aslandogan, Y.A.: Content-based image retrieval incorporating models of human perception. IEEE Trans. on Computers and Communication **2** (2004)
19. Frese, T., Boumann, C.A., Allenbach, J.P.: A methodology for designing image similarity metrics based on human visual system models. In: Proc. of SPIE/IS&T Conf. on Human Vision and Electronic Imaging II. (2002)
20. Chen, Y., Wang, J.Z.: Image categorization by learning and reasoning with regions. J. of Machine Learning Research (2004)
21. C.Carson, M.Thomas, S.Belongie, J.M.Hellerstein, J.Malik: Blobworld, A system for region-based image indexing and retrieval. In: Third Int. Conf. on Visual Information and Information Systems (VISUAL 99). (1999)
22. Wang, J.Z., Li, J., Wiederhold, G.: "simplicity: Semantics sensitive integrated matching for picture libraries ". IEEE Trans. Pattern Analysis and Machine Intelligence (2001)
23. Li, J., Wang, J.Z., Wiederhold, G.: IRM: Integrated Region Matching for Image Retrieval. In: Proc. of 8th ACM Int. Conf. on Multimedia, Los Angeles, California (2000) 147–156
24. Chen, Y., Wang, J.Z.: A region-based fuzzy feature matching approachto content based image retrieval. IEEE Trans. on Pattern Analysis and Machine Intelligence (2002)
25. Tubbs, J.D.: A Note on Binary Template Matching. Pattern Recognition **22** (1989) 359–365