RI99090 December 29, 1999 Subject: Computer Science

IBM Research Report

iPURE: Perceptual and User-friendly REtrieval of Images

Gaurav Aggarwal, Pradeep Dubey, Sugata Ghosal, Ashutosh Kulshreshtha, Abhinanda Sarkar
email: agaurav@in.ibm.com

IBM Research Division
IBM India Research Lab
Block I, I.I.T. Campus, Hauz Khas
New Delhi - 110016. INDIA.
Phone: +91-11-6861100, Fax: +91-11-6861555

iPURE: PERCEPTUAL AND USER-FRIENDLY RETRIEVAL OF IMAGES

Gaurav Aggarwal, Pradeep Dubey, Sugata Ghosal, Ashutosh Kulshreshtha and Abhinanda Sarkar

IBM India Research Lab
Block I, Indian Institute of Technology
Hauz Khas, New Delhi - 110016. India.
agaurav@in.ibm.com

ABSTRACT

Content-based Image Retrieval (CBIR) systems built around mathematical similarity models are often limited in their ability to capture user perception. On the other hand, relevance feedback approaches to CBIR in client-server environments incur a significant overhead in database search and image download time. A novel methodology for understanding user perception through a set of interactive tools is presented in this paper. The system uses a query by example methodology and uses color, texture and shape parameters for perceptual matching. It provides an object-level view of the query image through image segmentation. Traditional relevance feedback techniques on the searched results are supplemented with intra-query learning where the user perception is learnt through feedback on a set of images automatically generated from the query image itself, thus, alleviating the overheads of database search and image download. In addition, we allow the user to explicitly redefine the query point by manually modifying the query image. We demonstrate the feasibility and advantages of the approach with examples.

1. INTRODUCTION

A large number of content-based image retrieval systems have been built in the past few years. These system strive to search image databases using image centric features like color, texture, shape, position etc. of the query image rather than the traditional way of searching annotated databases using keywords. Most of these systems have focussed on some major issue of this multi-disciplinary problem. In order to point out the key differences, we present a high-level classification of the various systems in Figure 1. We point out that the distance of various systems from the origin does not have any special significance. The space is defined by three axes viz. Image Segmentation, Relevance Feedback and Modification of the query image.

Some CBIR systems, like Query by Image Content (QBIC), use mathematical similarity measures on global features and perform reasonably well for retrieving images containing *stuff*, i.e., scenes, textures etc. However, these systems do not perform as well on images of *things*, i.e., objects and general stock photography. Blobworld [2] and Netra [5] are two systems that use image segmentation to match an object-level view of images for retrieval. There are yet other global feature-based retrieval systems that incorporate relevance feedback techniques from the text retrieval domain to capture the user perception. These systems include MARS, PicHunter [3], SurfImage [6] and FourEyes among

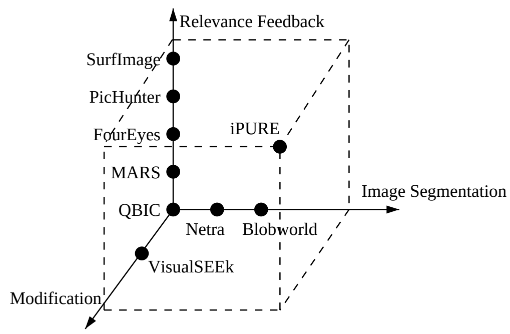

Figure 1: High-level classification of CBIR systems

many others. However, these relevance feedback based systems use global features only.

We propose a novel methodology which uses *automated modifications* of the query image itself and learning of user perception from intra-query learning on this set of modified images. Since the images are generated at the client-end itself, it saves the database search and image download time which is a significant overhead in the current relevance feedback approaches. VisualSEEk [7] only allows a user to manually modify the global features of the image while our technique modifies the retrieval parameters and generates synthetic images automatically. This makes the system more user-friendly since it requires the user to only mark the images as acceptable or unacceptable without requiring her to have any image processing expertise. User responses on this set of modified images is used to learn the initial weights of retrieval parameters. Furthermore, since our system uses image segmentation, it gives an object-level view of the image and hence, allows object-level modifications and retrieval of *things*. The system also employs relevance feedback techniques on the retrieved images using the weights learnt from feedback on the modified images. Thus, the iPURE system built using these techniques incorporates intra-query learning through query modification, in addition to relevance feedback on segmented images.

2. iPURE RETRIEVAL METHODOLOGY

The iPURE system is a segmentation-based image retrieval system that works in a client-server environment. The system architecture is shown in Figure 2. Database creation and updation is an off-line task where images are segmented using LUV color features. Feature-level segment descriptors as well as spatial descriptors are

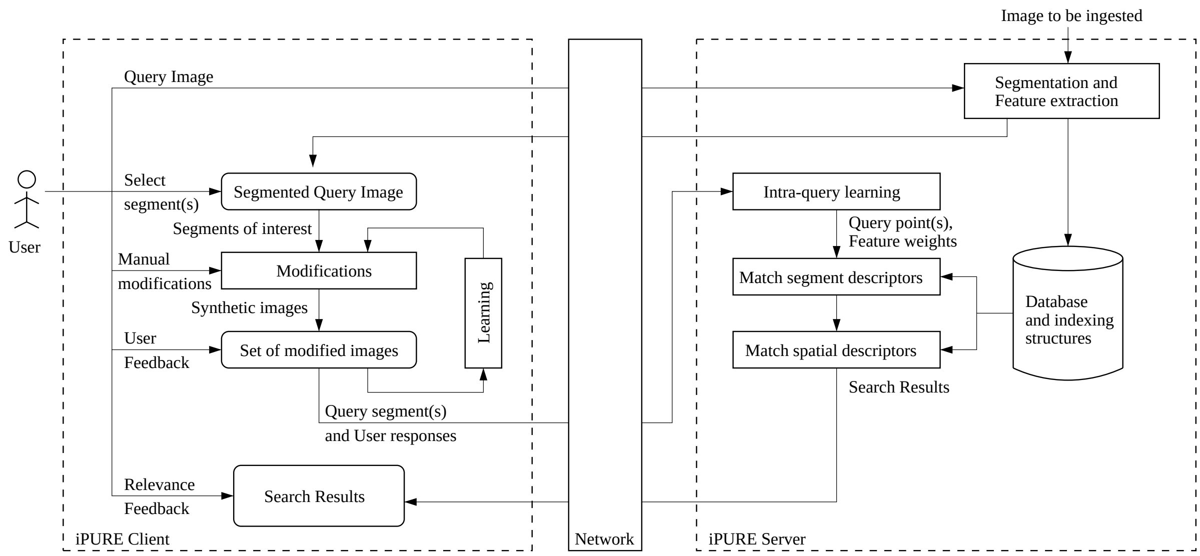

Figure 2: Architecture of the iPURE retrieval system

extracted for each segment in the image. These feature vectors are ingested along with the segmentation results in the database as shown in the top right corner of Figure 2.

The user starts the retrieval session with a query image which is transmitted to the server over the network. The server sends the segmentation results to the client. The user can select one or more segments of interest and can then either proceed to the retrieval process or go through the intra-query modification and learning before starting the retrieval process. Standard K-nearest neighbor matching using weighted Euclidean distance is used. The retrieval strategy for multiple-segment queries is similar to the ones used in Blobworld and VisualSEEk and proceeds in two steps. First, the best matches for individual segments are found using the segment descriptors. Then the topological inter-relationships amongst the retrieved segments are matched against those of the query segments using the spatial descriptors. A boundary around the matched segments is marked in the retrieved images so that the user may give feedback on a per-segment basis.

During intra-query modification, the client Java applet generates a set of modified images automatically. The user gives feedback on these modified images, i.e., whether the images are acceptable or unacceptable. Modification is performed essentially to generate training data for the learning module to compute both the relative weights of the segment descriptor components as well as compute the query point itself. We feel that such an interface is more user-friendly than asking the user to specify the relative importance of features like color, texture, position etc. as done in Blobworld and QBIC. An average user will be oblivious to these image centric terms. Furthermore, the iPURE system lets the user explicitly specify the query point which is specially useful when the user only has an approximate query to begin with.

3. INTRA-QUERY MODIFICATION AND LEARNING

The iPURE system distinguishes multiple-segment queries from the single-segment ones in order to better learn the user perception of complex queries [1]. When the user is interested in a single segment, the system generates perceptually significant modifications in the retrieval features for learning the relative importance of the features to the user. The set of basic modifications includes color change of the pixels in the segment (bit-flipping in the RGB domain), maximum horizontal and vertical translation of the segment within the image boundaries, scaling of segment about the centroid by 150% and 50%, and rotation about the centroid by 90°. If the user finds the image containing the color-flipped segment acceptable, it implies low relevance of color feature components for the user and subsequently the weights of the color components is lowered during retrieval. Relevance of other features is inferred similarly. In addition, the query point is redefined as the mean of feature vectors of acceptable images.

The iPURE system modifies the query shape itself to generate a variety of similar shapes (eight in the current implementation). These similar shapes are obtained by deforming a template generated by the query shape. We use the traditional Fourier descriptors to describe the shapes of the segments in an image. This template is then decomposed using a polar representation suggested in [8]. The three leading Fourier descriptors are used to obtain a smooth template that approximates the global features of the query shape. The template takes the form of an ellipse with the axes determined by the query shape. The modifications stretch and shrink these axes as well as increase and decrease the angle between them. Thus, these modifications are equivalent to stretching, shrinking and rotation of the global features of the shape. The local features captured by lower order Fourier descriptors are left unmodified and are simply added to reconstruct the modified shape. User's feedback on this set of shape modifications is used to compute the weights for the Fourier shape descriptors. This technique captures, e.g., deformations in a tree on a windy day where the global features deform the shape of the tree but the local features associated with the leaves on the branches are left unchanged.

The iPURE system uses the Wold model [4] for texture modification since it is both good for similarity based retrieval as well as for faithful reconstruction of the image from the modified model parameters. This model is based on 2-D Wold decomposition of homogeneous random field into the deterministic and purely-indeterministic fields. The Wold feature set is derived for the deterministic field which captures the periodicity and directionality in the texture. This feature set is the set of peak frequencies and their magnitude in Fourier domain. We recreate the texture from this feature set and the phase information, by taking the inverse Fourier transform of a Fourier field, where the Wold frequencies have the corresponding magnitude and phase while others are set to zero.

The iPURE system infers if the user is interested in texture at all, and if yes, then which Wold frequencies are important in her perception. The system does a two-step modification of the query to learn this. First, the segment texture is erased by applying an averaging filter. If the user finds the smoothed segment acceptable, it implies that the texture feature is not important. If the user finds this modification unacceptable, then it implies that the texture features should also be used for matching segments. The system then shows images with modified textures in the segment, recreated by removing one Wold frequency at a time. User's response on these modified images allows us to compute the weights given to different Wold frequencies in the adapted version of the similarity function given in [4].

There is often a *semantic* relationship between multiple segments of interest. Thus, when the user selects multiple segments, the iPURE system hypothesizes that the segments constitute one of the following: (i) multiple objects of interest, e.g., user selects an apple and an orange from the query image or (ii) a *composite semantic object*, e.g., when the user selects the different segments of a multi-color national flag, they actually form a composite entity, i.e., the flag or (iii) an object and a background, e.g., the sun and the sky [1]. Most other systems often treat multiple segments as multiple objects only. The iPURE system verifies these hypothesis by modifying the obvious semantic relationship, e.g., by breaking the contiguity of the multiple colored segments of a national flag. Rejection of the generated image by the user implies a semantic object and hence, the weight of composite shape features and topological relationships is increased during the second step of matching. Similarly, the enclosure relationship may be distorted by moving the sun outside the sky segment. Again, rejection of the generated image implies that the enclosure relationship should be verified during the second step of matching. Such inferences for multiple segments of interest substantially help in reducing the number of irrelevant images shown to the user.

4. USER-DRIVEN MODIFICATION

The iPURE system also provides a capability for manual modification of the query image for tuning the query point to better reflect her perception. The user can modify either the color, shape or texture features of the selected segment(s) and verify the visual effect of the changes made. We believe this interface is preferable over asking the user to sketch and color an image as is done in QBIC and VisualSEEk. Furthermore, segmentation allows the user to modify the properties at the object-level rather than global image-centric properties.

The iPURE system provides a color-grid using which the user can manually assign a color to the selected segment(s). When the user selects a color from the color-grid, the RGB value of each pixel in the selected segment is modified to make the average RGB of the selected segment same as the RGB value of the selected color from the color-grid. This operation preserves the color distribution in the segment(s).

The iPURE system also provides the user an ability to redefine the shape of the query object using a shape-grid. The shape-grid is a set of deformed object boundaries using the eight deformations of the template generated by the query shape. When the user picks one of these deformations from the shape-grid, the shape of the segment of interest is modified and the resulting image is shown to the user. The user may use this modified shape as the query shape.

5. EXPERIMENTS

We briefly present some results and refer to [1] for a detailed description and quantitative benefits of intra-query modification and learning. The current iPURE system uses the Corel stock photography database of 2200 images.

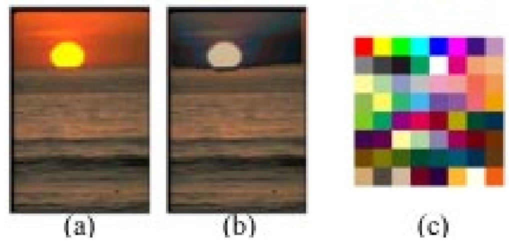

Figure 3: User-driven modifications of segments using a color-grid (a) Query image of sunset (b) Manually modified to moonlit sky (c) Color-grid

The benefit of user-driven modification is shown in Figure 3 where the user had an image of sunset as the approximate query image but wanted to retrieve images of moonlit sky. The user selected the yellow sun segment and changed the color to white by choosing the white color from the color grid. Similarly, the orange sky segment was changed to black. This modified image is then used for retrieving images of moonlit sky.

Figure 4: User-driven modifications of a segment using a shape-grid (a) Query image of highway sign (b) Arrow shape manually modified (c) Shape-grid

The use of a shape-grid to redefine the query shape is shown in Figure 4. The user selected the segment corresponding to the arrow sign in Figure 4(a) and then selected the modified shape from the shape-grid as shown in Figure 4(c). The modified image, shown in Figure 4(b), may then be used for retrieving images of distorted or broken highway signs.

An illustrative set of automatically modified images when the user selects the sun and the sky segment in the query image of Figure 3(a) is shown in Figure 5. The iPURE system checks the hypothesis that the sky segment should enclose the sun segment by

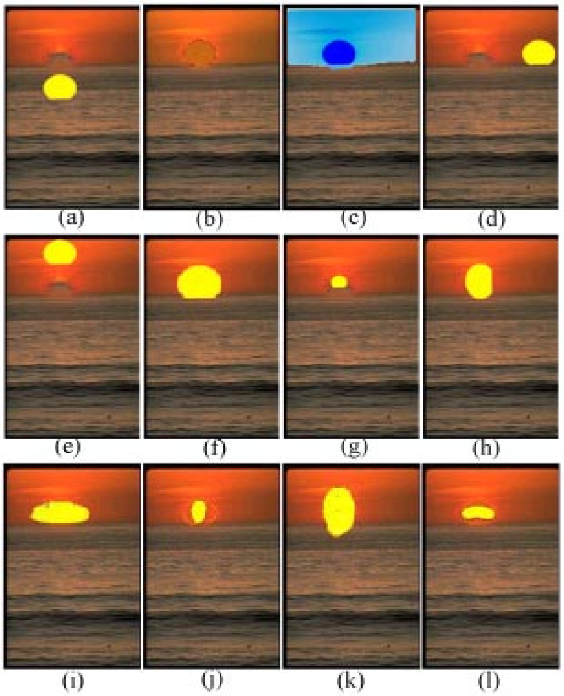

Figure 5: Automatic modifications for the sun and sky query segments: (a) Distorted enclosure (b) Average color (c) Color flipped (d) Horizontal translation (e) Vertical translation (f) Scaling-up (g) Scaling-down (h) Rotation (i–l) Shape modifications

showing an image in which the enclosure property is distorted, as shown in Figure 5(a). The user does not find this image acceptable and the iPURE system then proceeds to generate further modified images, ensuring that the sun is always within the sky. Additionally the sky segment is not translated, rotated, or scaled, since it is a background and such modifications would generate images that are meaningless in the visual sense and lead to incoherent user responses.

Further, we demonstrate the concept of the composite semantic object in Figure 6. The user selects the two segments corresponding to the black head and brown handle of the hammer in Figure 6(a). The system hypothesizes that since these two segments are contiguous, they might form a semantic object and hence proceeds to distort the contiguity as shown in Figure 6(b). The user rejects the generated image and the system learns the importance of contiguity and the composite shape of the semantic object. Further modifications treat the two segment as a single object and color change, translation, scaling, and rotation are done to both the segment simultaneously.

6. CONCLUSIONS

In this paper, we have proposed a novel client-server architecture-based iPURE system that performs perceptual and user-friendly content-based retrieval of images by automatically generating a set of modified query images at the client site. User perception is learnt based on the user feedback (in terms of relevance of images to him) on the set of modifications. The proposed system is intended for image databases containing "things", and employs image

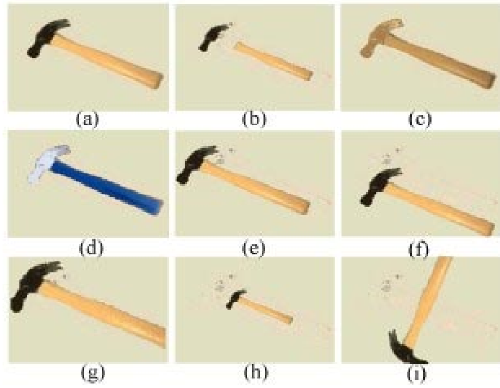

Figure 6: Automatic modifications for the hammer head and handle: (a) Query image (b) Distorted contiguity (c) Average color (d) Color flipped (e) Horizontal translation (f) Vertical translation (g) Scaling-up (h) Scaling-down (i) Rotation

segmentation for providing an object-level view of the query image to the user. This object-level view enables the user to manually redefine her query as well before she starts the actual retrieval process. Image segmentation also helps in pruning the dimensionality of the modification space. Currently, a set of generic heuristics is used for generating the automatic query modifications. It is expected that better retrieval performance can be achieved if the modifications can capture the distinct perceptual cluster attributes of the database population. A similar strategy has been advocated in PicHunter for specifying a set of query images to the user at the search or browse time [3].

7. REFERENCES

[1] G. Aggarwal, S. Ghosal, and P. Dubey. Efficient Query Modification for Image Retrieval. In *Proc. of IEEE CVPR'2000*, South Carolina, U.S.A., 2000.

[2] C. Carson, S. Belongie, H. Greenspan, et al. Region-based image querying. In *Proc. of IEEE CBAIVL'97*, San Juan, Puerto Rico, 1997.

[3] I.J. Cox, M.L. Miller, T.P. Minka, et al. An optimized interaction strategy for Bayesian relevance feedback. In *Proc. of IEEE CVPR'98*, Santa Barbara, U.S.A., 1998.

[4] F. Liu, and R.W. Picard. Periodicity, directionality and randomness: Wold features for image modelling and retrieval. *IEEE Trans. on PAMI*, 18(7):722–733, July 1996.

[5] M.Y. Ma, and B.S. Manjunath. NETRA: A toolbox for navigating large image databases. *Multimedia Systems*, 7(3):184–198, 1999.

[6] C. Nastar, M. Mitschke, and C.Meilhac. Efficient query refinement for image retrieval. In *Proc. of IEEE CVPR'98*, Santa Barbara, U.S.A., 1998.

[7] J.R. Smith, and S.F. Chang. VisualSEEk: A fully automated content-based image query system. In *Proc. of ACM Multimedia*, Boston, U.S.A., 1996.

[8] R.K.K. Yip, P.K.S. Tam, and D.N.K. Leung. Applications of elliptic Fourier descriptors to symmetry detection under parallel projection. *IEEE Trans. on PAMI*, 16:277–286, 1994.