Segmentation Tree based Multiple Object Image Retrieval

Wei-Bang Chen
Department of Mathematics and Computer Science
Virginia State University
Petersburg, VA 23806, USA
wchen@vsu.edu

Chengcui Zhang and Song Gao
Department of Computer and Information Sciences
The University of Alabama at Birmingham
Birmingham, AL 35294, USA
zhang@cis.uab.edu, gaos@uab.edu

Abstract— **Inaccurate image segmentation often has a negative impact on object-based image retrieval. Researchers have attempted to alleviate this problem by using hierarchical image representation. However, these attempts suffer from the inefficiency in building the hierarchical image representation and the high computational complexity in matching two hierarchically represented images. Existing approaches construct the hierarchical image representation in two steps. The first step is to perform segmentation at different image resolutions, and the second step is to construct a hierarchical representation of the image by associating segments from different resolutions. In this research, an innovative all-in-one-run approach is proposed that concurrently performs image segmentation and hierarchical tree construction, producing a hierarchical region tree to represent the image. In addition, an efficient hierarchical region tree matching algorithm is proposed with a reasonably low time complexity and used in multiple object image retrieval. The experimental results demonstrate the efficacy and efficiency of the proposed approach.**

Keywords-content-based image retrieval; multi-object retrieval; hierarchical region-tree; multi-resolution image segmentation

I. INTRODUCTION

As digital imaging has emerged from its infancy in the past decade, more and more digital images have become available. As the adage suggests, "a picture is worth a thousand words." Information embedded in an image usually provides a more clear and succinct way to present an idea than a substantial amount of text. The emerging needs in retrieving information from images brings researchers' attention, and thus, image retrieval has been an extremely active research area in the past decade. Many efforts have been made to address this challenging issue. These efforts can be classified into two categories: (1) text-based image search, and (2) content-based image retrieval (CBIR).

In most conventional text-based image search systems, all images in the search scope must first be annotated. The annotations such as the file name, caption, keywords, tags, and other text-based descriptions, are stored in the associated metadata. Then, the text-based database management systems (DBMS) retrieve images based on the annotations stored in the associated metadata [1]. The major problems of text-based image retrieval systems are: (1) they heavily rely on image annotations or surrounding text rather than semantic content, and thus, cannot distinguish homonyms; (2) it would be difficult to precisely describe all visual content in an image with a limited set of words [1], and the perception and interpretation of visual content varies from person to person.

In contrast to text-based image search, content-based image retrieval (CBIR) has been introduced to cope with the issues that arise in text-based image retrieval systems. CBIR systems search images based on the visual content of images. The concept of CBIR was first introduced by Kato in 1992 to describe the automatic process of retrieving images from an image database according to the visual features extracted from images [2]. CBIR systems view the query image and all target images in the database as a collection of primitive visual features such as color, texture, shape, and spatial location. On the basis of these primitive visual features, CBIR systems measure the similarity between a query image and each target image in the database. Then, the target images are ranked in the decreasing order of their similarities to the query image [3]. From this image retrieval process, three fundamental bases can be summarized for content-based image retrieval framework, namely primitive visual feature extraction; multi-dimensional indexing; and retrieval system design [4].

Content based image retrieval systems can be further categorized into two major approaches, including full image search and object-based image retrieval. The full image search retrieves images based on the global visual features extracted from the whole image [5]. In contrast to full image search, another line of approaches is object-based image retrieval which attempts to capture the high level concepts embedded in images such as objects. To perform object-based search, it is essential to extract embedded objects in images with image segmentation which is known to be one of the most challenging issues in the field of image processing.

In image segmentation, one of the challenging issues is over- and under-segmentation. Due to the imperfection of the segmentation algorithms, segmentation results obtained from most of the existing segmentation algorithms are often over-segmented and/or under-segmented. While both over- and under-segmentation cause problems, under segmentation has a bigger negative impact on object-based image retrieval. The reason is that an under-segmented region represents several different objects with one region in the image, which is less useful in the object-based image retrieval. On the other hand, an over-segmented region could still represent

978-0-7695-4875-3/12 $26.00 © 2012 IEEE
DOI 10.1109/ISM.2012.49

part of an object. Therefore, most existing segmentation algorithms tend to over-segment. Thus, the main challenge is how to alleviate the problem of over segmentation in object-based image retrieval.

While those high level concepts of users' interests come naturally to a human being, they pose a big challenge to computer systems due to the so-called semantic gap. This is because computer systems can only recognize those low level primitive visual features, but not the high level concepts. In order to bridge the semantic gap, Li et al. introduced the integrated region matching (IRM) scheme which measures the overall similarity between images according to the overall similarity between two sets of image segments [6]. Another approach, dynamic region matching (DRM) was also introduced in 2008 by Ji et al. to address the same issue [7]. In 2010, Zhang et al.'s works have demonstrated that combining integrated region matching (IRM) scheme, a measure for the overall similarity between images according to the similarity between two sets of image segments, with users' relevance feedback (RF) in a multi-object based image retrieval framework makes the automatic discovery of user desired objects possible [8]. However, the abovementioned approaches suffer greatly from inaccurate segmentation especially over-segmentation. Further, these approaches retrieve "objects" on the basis of a collection of independent segments/regions which may not individually correspond to semantic objects, without considering the associative relationships between image segments. On top of that, the adverse effect of inaccurate segmentation has become a major bottleneck that impedes the progress of object-based image retrieval systems. For example, over-segmented regions that originate from different objects may be extremely similar, and thus, may aggravate the problem of false positives. We believe the key to alleviating the above issue is a new systematic and hierarchical representation of visual information, and the corresponding analysis and retrieval framework that make it possible for a machine to interpret an image in terms of its containing regions and their relationships. For this reason, it is essential to preserve the spatial and neighboring relationships between and among segments in order to model the image content. One possible solution is to use hierarchical image representation to preserve such relationships between and among segments.

Existing approaches construct a hierarchical image representation in two steps [9-14]. The first step is to perform segmentation at different image resolutions, and the second step is to construct the hierarchical representation of the image by associating segments from different resolutions. This two-step process has low efficiency due to high time-complexity associated with the multi-scale image analysis.

The goal of this research is to develop an effective and efficient multiple-object image retrieval framework which can alleviate the over-segmentation problem by introducing the hierarchical image representation, but does not suffer from the inefficiency during the construction of the image hierarchy and the comparison of hierarchical representations of images. In this paper, we introduce a multiple-object image retrieval system named (MOIR) in order to achieve the above goals. In the proposed MOIR framework, we develop an efficient algorithm named "Multi-Resolution Image Analysis" (MRIA) to perform image segmentation and construct the image hierarchy all in one run. This is achieved by designing a branch-and-bound-like algorithm that performs image segmentation and hierarchical tree construction concurrently, and the analysis progresses from low resolution to higher resolution and uses certain constraints to enhance performance. In addition, we also design an efficient algorithm that is used to compare two image hierarchies representing two images.

The remainder of this paper is organized as follows. In Section 2, we describe the details of the proposed multiple-object image retrieval framework. The experimental results are demonstrated in Section 3. Section 4 concludes this paper.

II. PROPOSED METHOD

A. Framework Overview

The high level architecture of the proposed multiple-object image retrieval (MOIR) framework is illustrated in Figure 1.

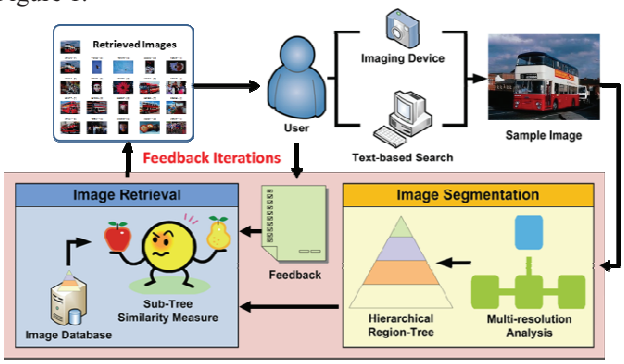


Figure 1. The proposed multiple-object image retrieval framework.

In brief, the proposed multiple-object image retrieval process starts with the submission of a query sample image. Then, the proposed multi-resolution image analysis (MRIA) is performed on the query image to build a hierarchical region tree. In the next step, the proposed MOIR framework measures the similarity between the query image and each target image in the database, which is achieved by the comparison of two hierarchical region trees, representing the query image and target image, respectively. Then, target images are ranked in the decreasing order of their similarities to the query image. According to the top 20 images in the ranked list, users provide feedback to the retrieval system in order to refine the retrieval results.

B. Hierarchical Image Representation

As aforementioned, the main challenging issue in object-based image retrieval is how to alleviate the problem of over segmentation due to most existing segmentation algorithms tend to over-segment an image. Researchers introduce hierarchical image representation to preserve the relationship between and among segments [15-17] in order to reduce the negative impact of inaccurate segmentation in object-based image retrieval. The hierarchical image representation is a

flexible and convenient way to mirror the multi-scale processing in the human visual system.

The conventional approach to hierarchical image representation is a two-step process, including image segmentation followed by region tree construction. The first step is to perform segmentation on images presented in different resolutions from the highest (the original image) to the lowest, producing a segmentation mask for each resolution. The second step is to construct the hierarchical representation of the image, i.e., a region tree, by associating segments from different resolutions. However, these multi-level analysis approaches suffer from a high computational complexity. First, performing a full-scale segmentation at each different image resolution is itself complicated enough, let alone the need of one extra run through all resolutions to associate segmented regions. One of our goals in this research is to design a novel hierarchical image segmentation algorithm that possesses the following characteristics: (1) preserving the spatial relationships between and among segmented regions as a hierarchical region tree to represent an image; (2) performing image segmentation and hierarchical region tree construction in a concurrent manner to reduce the computational complexity; (3) including an branch-and-bound-like algorithm that performs image analysis from low resolution to higher resolution in order to mitigate the inefficiency during the multi-level analysis.

In this paper, we proposed a multi-resolution image analysis (MRIA) algorithm that performs hierarchical image segmentation with the above desired characteristics. The proposed MRIA algorithm is inspired by the human visual system. Imagining you are standing on an open field and a red sports car is moving toward you at a very far distance. Initially, your eyes can only see a tiny red object without any detail due to the visual acuity of the visual system. When the tiny red object is moving closer, your visual system has the ability to recognize the object as a red sports car but still cannot capture fine details of the car. Later, when the car approaches close enough, your eyes can distinguish fine details of the car such as the vehicle brand logo and the textures of wheels.

The above observation indicates that our visual system has limited resolving power and our brain only recognizes an object when our visual system provides enough details, the combinatorial of various primitive visual features, about the object being observed. This phenomenon also implies that when an object is located at a far distance, our visual system can only perceive down-sampled signals from the object. In other words, human visual system cannot provide enough details about that object until the sampling rate reaches certain level. The entire process reflects that human brain actually performs a multi-resolution analysis through our visual system, which motivates us to adopt a similar multi-level analysis process into the proposed MRIA algorithm.

In signal processing, down-sampling is known to be a process that removes bandwidth in high-frequency and preserves bandwidth in low-frequency in data. Therefore, the most prominent regions in images can be obtained even with low sampling rate, while the detailed information can be revealed at higher sampling rates. In general, the most prominent regions in images usually indicate either backgrounds or a target object in close-up shot. If we progressively increase the sampling rate, more and more details will be become evident for each prominent region discovered previously. The process of gradually increasing the sampling rate naturally forms a region-based hierarchical tree with different levels of details. Moreover, as an added benefit, performing analysis on low resolution images is much more efficient than that of the high resolution images.

The flowchart of the proposed MRIA algorithm is depicted in Figure 2. As aforementioned, one of our goals in this research is to mitigate the inefficiency of the multi-level analysis. To achieve this goal, our approach is to design a branch-and-bound-like algorithm that performs image analysis from the lowest resolution and progresses to higher resolutions if necessary. Apparently, the analysis of a low resolution image is much faster than that of a high resolution image. Therefore, the first step in the proposed algorithm is to obtain down-sampled images. Discrete wavelet transformation (DWT) is known to be an efficient method to transform the original image into a series of down-sampled images. In this paper, Haar wavelet transform is used to produce a series of down-sampled images by reducing image size by half in each dimension [18] each time it gets down-sampled. In addition, a minimal image size of 8-by-8 pixels is preset as a constraint because any image smaller than this preset size will be too coarse to differentiate meaningful objects.

The second step starts with creating a root node to represent the entire image at the lowest resolution and extracting primitive visual features from the lowest resolution image. Any visual features that are suitable for segmentation and robust during down-sampling can be readily used in this framework, though our focus is not to explore the best features for segmentation or image retrieval. For this reason, we adopt MPEG-7 dominant color descriptor as its primitive visual features in this paper [19].

By considering the entire image as one region, the first level image segmentation (region growing algorithm [20] is used in this paper) is performed on the region, producing a segmentation mask. Then, we increase the image resolution and up-scale the segmentation mask in order to obtain more details about each segmented region. For each segmented region, a child node is created to represent that region, and primitive visual features are extracted from the region, followed by the second level image segmentation on each region produced by the first level segmentation. With more details revealed for a region at each higher resolution, the subsequent higher-level segmentation plays an important role in determining the homogeneity of the region. More specifically, a region is considered homogeneous if no new segment can be segregated from that region, indicating that there is no need to further process this region in the subsequent analysis. On the other hand, for a region that is not sufficiently homogeneous will likely to be further segmented into smaller segments at a higher image resolution, and a new segmentation mask is produced for that region. This process will continue until either of the following criteria is satisfied. The two stopping criteria are:

(1) no region can be further segmented, or (2) the highest image resolution is reached.

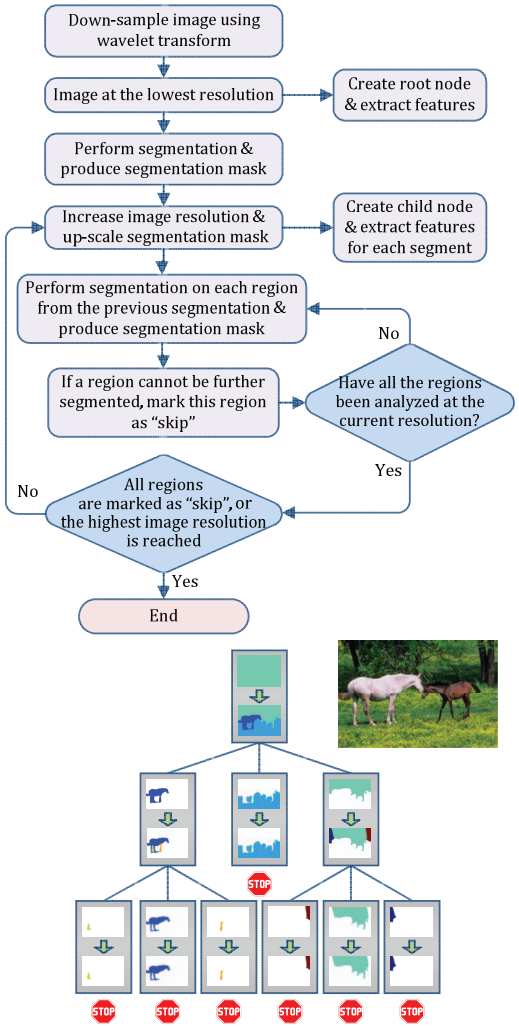


Figure 2. The proposed multi-resolution image analysis (MRIA) framework for hierarchical image representation.

In summary, the proposed MRIA algorithm first segments an image at the lowest resolution, and performs subsequent segmentation for each previously generated region only when necessary, i.e., when that region is not sufficiently homogeneous. During the same process, a hierarchical tree representation is constructed (in a top-down manner) along with the multi-resolution segmentation results. The key point in this process is to avoid unnecessary image segmentation at any higher image resolution – if a sub-tree/branch, which represents a region in the image, is considered homogenous, it will be removed from any subsequent segmentation (pruning of the analysis space). In this way, the computational cost can be dramatically reduced.

Although we can use the image hierarchy to preserve the associative relations between and among segments, the negative impact of over-segmentation still remains unsolved for object-based image retrieval until we make use of the hierarchical tree matching in the image retrieval process. In the next step, we utilize the preserved associative relations to alleviate the over-segmentation problem by introducing the hierarchical region tree matching.

C. Hierarchical Segmentation Tree Matching

With the proposed MRIA algorithm, the query image and all target images in the database are segmented into regions at different resolutions. For each image, the relations among those segmented regions are concurrently preserved in the form of a hierarchical region tree. As aforementioned, an image hierarchy reflects that image's visual composition, and thus, provides a way to model the visual content of that image. Figure 3 provides some examples of hierarchical region trees. A typical hierarchical tree consists of three types of nodes including one root node (R), leaf nodes (L), and inner nodes (I). The root node represents the entire image as a single region. A leaf node represents a region with consistent visual features and cannot be further partitioned into sub-regions in that feature space. An inner node represents a region that consists of at least two sub-regions. In this paper, we refer to a sub-tree of a tree T as a tree consisting of a node and all of its descendants in T. Thus, the sub-tree corresponding to the root node is the entire tree; the sub-tree corresponding to any inner node (I) in T is defined as a proper sub-tree (P). For each proper sub-tree (P) or leaf (L) in a hierarchical image representation, it can represent multiple objects, a single object, or part of an object.

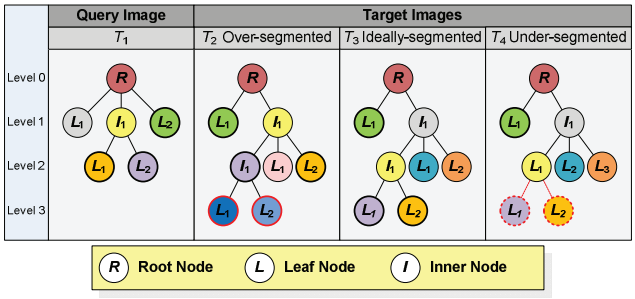


Figure 3. Four examples of hierarchical region trees.

Figure 3 demonstrates four hierarchical region trees T_1, T_2, T_3 and T_4 which model the content of a query image (T_1) and three target images (T_2, T_3 and T_4) in the database, respectively. As shown in Figure 3, symbols R, L, and I represent the root node, a leaf node, and an inner node, respectively. Numbers in the subscripts indicate the ordinal value of a specific type of node (L or I), at that level. The numbering of ordinal values restarts from 1 at each new level. The corresponding regions from different hierarchical region trees, i.e., from different images, have the same color.

Traditional CBIR frameworks, such as SIMPLIcity, measure object relevance on the basis of the comparison of two sets of objects which does not consider the relationships among segments in an image [21]. Unlike the conventional CBIR frameworks, using hierarchical region tree in the proposed object-based CBIR system provides additional information on the relationships among the segments in an

image and is expected to reduce the negative impact of inaccurate segmentation, especially over-segmentation. Taking into account the relationships along with individual image segments allows the proposed CBIR framework to better measure the similarity between two regions (not necessarily the regions corresponding to leaf nodes) from two images. This idea transforms the object comparison problem into proper sub-tree comparison.

As aforementioned, image segmentation is an extremely difficult problem. Although an object may be ideally-segmented, quite often an object suffers from over-segmentation or under-segmentation problems. An ideally-segmented object corresponds to a leaf node in a tree, but a leaf node may represent an under-segmented region which contains two or more objects. An over-segmented object corresponds to an inner node in a tree. Figure 3 depicts an over-segmented object in T_2, an ideally-segmented object in T_3, and an under-segmented region in T_4.

For convenience, we will use a shorthand representation to refer to a node in a tree throughout the rest of this paper. The shorthand representation is defined as:

(Tree #, Level #, Node_Type.Ordinal_Value)

The root node is at Level 0. For example, when we refer to the inner node (I_1) located at Level 2 in tree T_2, the shorthand representation of this node is (2, 2, I.1).

In Figure 3, as indicated by the same color, (2, 2, I.1) in T_2 corresponds to the same object ideally segmented in T_1 (1, 2, L.2) and T_3 (3, 3, L.1), but is further partitioned into (2, 3, L.1) and (2, 3, L.2) in T_2. This indicates that this object in T_2 is over-segmented. (1, 2, L.1) and (1, 2, L.2) represent two ideally segmented objects in tree T_1. However, the corresponding nodes do not exist in T_4. This is because that the node (4, 2, L.1) in T_4, which should correspond to the node (1, 1, I.1) in T_1, is under-segmented. In other words, two objects (1, 2, L.1) and (1, 2, L.2) are both included in one region (4, 2, L.1) in T_4 but they cannot be separated by segmentation on that image. For illustration purposes, we depict these two nodes from T_1 in T_4 with red dotted circles as (4, 3, L.1) and (4, 3, L.2), though they don't exist in T_4. Although the nodes (1, 2, L.1) and (1, 2, L.2) probably cannot be matched with any node in T_4, their parent node (1, 1, I.1) can still be matched to (4, 2, L.1).

From the above examples, three types of comparison can be concluded, including leaf to leaf (L-L) comparison, leaf to sub-tree or sub-tree to leaf (L-P/P-L) comparison, and sub-tree to sub-tree (P-P) comparison. The above three types of comparisons are actually performed through measuring the similarity between their primitive visual features. The L-L comparison measures the similarity between two segments which correspond to two leaf nodes. The L-P/P-L comparison simply measures the similarity between a segment that corresponds to a leaf node and a set of segments that correspond to a sub-tree. The P-P comparison calculates the similarity between two sets of segments that correspond to two sub-trees, respectively.

We expect that the similarity measure derived from the above three types of comparisons can reduce the negative impact of over-segmentation. This is because when matching two objects that either or both are over-segmented, the optimal object matching can still be achieved through a L-P/P-L or P-P comparison. However, we are not very optimistic about using hierarchical region trees to alleviate the problem of under-segmentation. Our take on this is that most existing image segmentation algorithms, especially those used in object-based image retrieval systems, tend to over-segment an image so that the retrieval performance is largely affected by over-segmentation [22]. Thus, we argue that by alleviating the problem of over-segmentation, the state-of-the-art of multiple object image retrieval can be advanced. In this research, we make sure that the proposed hierarchical image segmentation algorithm tends to over-segment an image but bounded by an acceptable rate of such.

A performance issue in terms of efficiency also emerges from the aforementioned comparisons. This is because there could be many sub-trees in one hierarchical region tree, not to mention when comparing all proper sub-tree pairs from a given pair of trees. For this reason, an efficient algorithm for matching two hierarchical region trees is developed in this paper. In order to avoid excessive computational cost in proper sub-tree comparison, our idea is to calculate the sub-tree similarity based on previously calculated similarity values during subsub-tree comparison, similar to the idea of dynamic programming. We use the following example to explain the proposed segmentation tree matching algorithm.

Figure 4 exemplifies two hierarchical region trees – *A* and *B*, representing a query image (*A*) and a target image (*B*), respectively. In matching two hierarchical region trees, our goal is to find, for each node in *A*, its best matching node in *B*. Recall that when building a region tree, all nodes are created in the order of top-to-down and left-to-right. In order to reuse the previously calculated similarity values, the tree comparison is performed in the reverse order. The comparison starts from matching the leaf node (*A*7) in *A* with each node in *B*. In this round of matching, there are 3 L-L comparisons, i.e., *A*7-*B*5, *A*7-*B*4, and *A*7-*B*3. After that, there are 2 L-P comparisons, i.e., *A*7-*B*2 and *A*7-*B*1. When performing a L-P comparison, the similarity between a leaf node and a sub-tree is defined as the highest similarity between the leaf node and a node in the sub-tree (including the root node of that sub-tree). However, there is no need to match the leaf node in the query image with every child node in that sub-tree of *B*. In fact, according to our reverse order of comparison, the similarity between that leaf node in *A* and every child node in the sub-tree of *B* has been previously calculated. Following the same process, the comparison continues and at a later time reaches the matching of an inner node (*A*3) with each node in *B*. In this round of matching, there are 3 P-L comparisons, i.e., *A*3-*B*5, *A*3-*B*4, and *A*3-*B*3. In addition, there are 2 P-P comparisons, i.e., *A*3-*B*2 and *A*3-*B*1. In each P-L comparison involved in this step, since the similarity between each child node of *A*3 and each leaf node of *B* has been calculated already during previous steps, there is no need to calculate them again, and the only additional computation incurred is the calculation of similarity between *A*3 itself and that leaf node in *B*. When comparing two proper sub-trees such as *A*3-*B*2, we first measure the inner node similarity (*INS*) which is defined as the similarity

between the two root nodes of two sub-trees. If the inner node similarity exceeds a predefined threshold value (> 90% similarity in our case), we further measure the highest child node similarity (*CNS*) between the two sub-trees. It is worth noting that the *CNS* can be directly derived from the child node similarity scores calculated in previous steps. The proper sub-tree similarity (*PSS*) is defined as the maximum of *INS* and *CNS* as formalized in the following equation.

$$PSS = \max\{INS, CNS\}$$

Assume there are m target images in the database. The similarity value between the query image and each target image can be efficiently measured using the proposed hierarchical region tree comparison algorithm, resulting in a vector of length n, where n is the number of nodes in the query image. The collection of aforementioned vectors forms a matrix of size m-by-n, and we name it the node similarity matrix. Each row in the matrix represents a target image, and each column heading in the matrix corresponds to a node in the query image. An entry $[m_i, n_j]$ records the highest similarity value between the n_j^{th} node in the query image and all the nodes in the hierarchical tree of the m_i^{th} image. According to the similarity scores stored in the node similarity matrix, the proposed multiple-object image retrieval framework can obtain the overall similarity by calculating the row sum, returning a ranked list of images to the user as the initial retrieval results. In addition, the node similarity matrix is used in the subsequent user relevance feedback process which progressively discovers the object(s) of the user's interests.

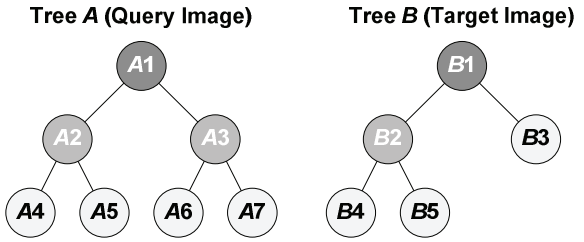


Figure 4. Matching two hierarchical region trees

D. Relevance Feedback

In addition to the development of an efficient sub-tree similarity measure, another challenge remained in the domain is how to discover the objects of the user's interest given the user's scarce and imbalanced feedback information as training data. We also want to avoid the proper sub-tree comparison during feedback iterations due to the expensive computational cost of sub-tree matching. For these reasons, our idea is to build a classifier that makes the maximum use of users' relevance feedback, learns user-desired object(s) from the node similarity matrix and user feedback, and refines the retrieval results.

To achieve this goal, the first step is to collect the user's feedback on the retrieval results. As aforementioned, the proposed MOIR framework calculates the row sum from the node similarity matrix which represents the overall similarity between the query image and each target image. The MOIR system ranks the target images in the descending order of their similarity to the query image, and returns the top 20 images (as the initial results) to the user for feedback. The user then provides feedback on those 20 images by giving either a positive or a negative label. A positive label is given if and only if the image containing all objects of the user's interest. Otherwise, a negative label is provided. The user's feedback is then used by the retrieval system to learn his object(s) of interest. Since only 20 images are returned to the user for feedback, the amount of feedback information is scarce in nature and can be extremely imbalanced (e.g., only 2~3 images are positive among the top 20). However, returning more images for user feedback could bring a big burden to the user.

The second step is to associate the user's feedback with the node similarity matrix. Recall that in the node similarity matrix, each column heading represents a node in the query image, and each row represents a target image. Since the user-desired object(s) must exist in the query image, one or more columns in the node similarity matrix represent the object(s) of the user's interest. It is not a trivial task to directly identify the relevant column(s), i.e., relevant object(s), in the node similarity matrix due to scarce feedback information. Instead of directly identifying the relevant column(s), we use one-class support vector machine (SVM) [23] to build a classifier and let the classifier determine the importance of each column/object in the query image. The idea is that we consider each row in the node similarity matrix as a feature vector used for SVM training, representing the similarity between the query image and a target image in term of object similarity. Further, we use the user's feedback on each returned top target image as a class label. All positive samples belong to one class which represents relevant images while all negative samples belong to another class which represents irrelevant images. Then, a set of distinct target images with the user's feedback are cumulatively collected as training samples through each feedback iteration. The training samples are fed to the one-class SVM to train the classifier. This trained classifier is then used to test the relevance of all target images in the database and rank them according to their decision values generated from the SVM classifier. In this way, we can progressively refine the retrieval results by maximizing the usage of all of the user's feedbacks collected through multiple iterations without sacrificing the efficiency because there is no need to recalculate the node similarity matrix.

III. Experimental Results

A. Dataset Description

The experiments in this paper are performed on a dataset containing 10,000 images collected from Corel Image Database. Unlike the traditional way where Corel category labels are used as ground-truth, we procure our own ground-truth for evaluation. Specifically, we define 50 objects and manually annotate images containing these objects. Many of these objects (e.g., blue sky, red car, and roadway) occur in several Corel categories. Our ground-truth labels are those manually annotated objects instead of Corel category labels.

B. Multi-Resolution Image Analysis (MRIA) Assessment

The performance of the proposed MRIA algorithm is evaluated through two experiments, including the efficiency analysis and the efficacy analysis.

1) MRIA efficiency analysis: In this experiment, we objectively assessed the performance of the proposed MRIA algorithm in terms of segmentation efficiency that quantifies the efficiency of a segmentation algorithm on the basis of the total number of pixels analyzed in the algorithm. The segmentation efficiency of an algorithm A applied to an image I is defined and formalized in the following equation.

$$\text{Segmentation Efficiency}(A, I) = 1 - \frac{\sum_{i=1}^{n_I} \#\text{ of pixels analyzed at level } i}{\#\text{ of pixels in the original image}}$$

where i represents the level in the multi-resolution image pyramid, and i=1 indicates the lowest image resolution in the image pyramid; n_I is the level of the highest image resolution processed for image I. Based on our experiment on the 10,000 images, the average segmentation efficiency of the proposed MRIA algorithm is 98.26%. This indicates that our approach is very efficient in segmenting an image and constructing the hierarchical image representation in one run.

2) MRIA efficacy analysis: We introduce a subjective quality assessment experiment to evaluate the efficacy of the proposed MRIA algorithm in terms of image segmentation quality. This experiment compares the segmentation results of the proposed MRIA algorithm and a hill-climbing based color k-means segmentation algorithm (HCK) [24, 25]. To ensure the integrity of subjective evaluation, 9 evaluators perform a blind review through a web interface. The evaluators vote the best segmented image from the two displayed segmented images produced by our algorithm and HCK, respectively. The evaluation system also provides a neutral option, if both segmented images are comparable. The results of the assessment are presented in Table I.

TABLE I. SUBJECTIVE SEGMENTATION QUALITY ASSESSMENT

	Comparable	MRIA is better	HCK is better
MRIA vs. HCK8	**20%**	**67%**	**13%**
MRIA vs. HCK10	**19%**	**73%**	**8%**

In Table I, the numbers '8' and '10' represent the different number of bins used in the HCK algorithm. Table I demonstrates that the image segmentation quality of the proposed MRIA algorithm significantly outperforms the HCK algorithm.

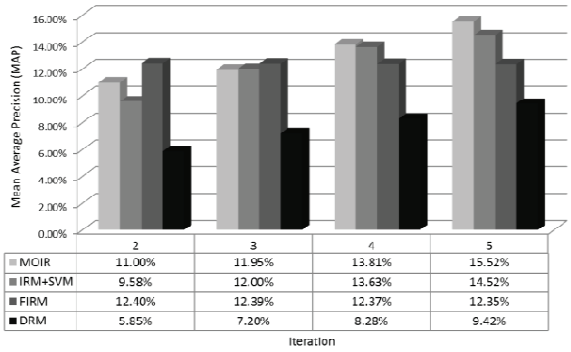
C. Multiple Object Image Retrieval (MOIR) Assessment

Two commonly used standard evaluation metrics, Average precision (AP) and mean average precision (MAP), are used in the subsequent experiments in order to fairly compare the proposed MOIR framework to other existing frameworks. We choose these two measures because they not only simultaneously take into account precision, recall, and rank, but also have been shown to have exceptionally good discrimination power and robustness.

Based on the MAP, we compare the proposed MOIR framework to three state-of-the-art frameworks, including integrated region matching (IRM) [6], feedback-based image clustering and retrieval framework (FIRM) [8], and dynamic region matching (DRM) [7]. In order to make this experiment a fair comparison, SVM is integrated into IRM for learning the user's feedback since IRM itself is designed for matching two sets of segments but without the ability to incorporate the user's relevance feedback.

1) Single object image retrieval: In the first experiment, we demonstrate the effectiveness of the proposed MOIR framework in single object retrieval based on 560 query images from 11 categories including dinosaur, red bus, pyramid, white rabbit, bullet, yellow car, yellow bus, bonsai, tiger, penguin, and shoji. Figure 5 shows the retrieval performance of each framework in terms of MAP.

From Figure 5, it can be observed that after 4 feedback iterations, the MAP value of the proposed MOIR framework reaches 15.52%, which is 1%, 3.17%, and 6.1% higher than that of the IRM+SVM, FIRM, and DRM, respectively. The proposed MOIR framework significantly outperforms other frameworks since the number of queries (560 queries) and the size of retrieval scope (10,000 target images) are large enough for us to claim that 1% increase in MAP value is significant. Further, the MAP value steadily increases through the feedback iterations, which indicates the robustness and effectiveness of the relevance feedback.



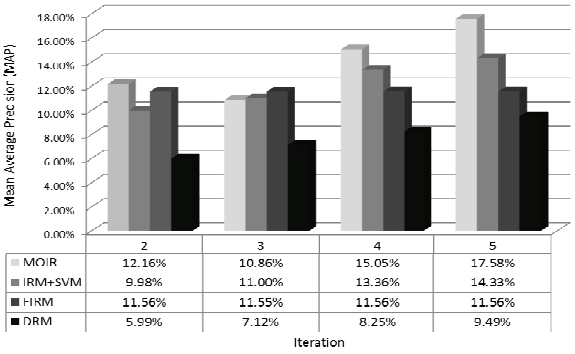
	2	3	4	5
MOIR	11.00%	11.95%	13.81%	15.52%
IRM+SVM	9.58%	12.00%	13.63%	14.52%
FIRM	12.10%	12.39%	12.37%	12.35%
DRM	5.85%	7.20%	8.28%	9.42%

Figure 5. Single object retrieval results (560 queries)

2) Multiple object image retrieval: In this experiment, we demonstrate the effectiveness of the MOIR framework in multiple-object retrieval based on 201 query images from 9 different query object combinations, including: bonsai + shoji (18), blue sky + red bus (34), pyramid + blue sky (21), white rabbit + snow (11), gun + bullet (19), red airplane + blue sky (18), red car + roadway (45), yellow bus + roadway (19), yellow car + roadway (16) where the number in the parentheses represents the number of query images in that combination. The performance for each framework is shown in Figure 6.

Figure 6 shows that after four feedback iterations, the proposed MOIR framework significantly outperforms IRM+SVM, FIRM, and DRM by 3.25%, 6.02%, and 8.09%, respectively. Similarly, the MAP value of our algorithm also

increases through the feedback iterations, again indicating the effectiveness and robustness of the relevance feedback.



	2	3	4	5
MOIR	12.16%	10.86%	15.05%	17.58%
IRM+SVM	9.98%	11.00%	13.36%	14.33%
HRM	11.56%	11.55%	11.56%	11.56%
DRM	5.99%	7.12%	8.25%	9.49%

Figure 6. Multiple-object retrieval results (201 queries)

IV. CONCLUSIONS

This paper presents an innovative human-centered multiple-object image retrieval framework which seamlessly integrates a multi-resolution hierarchical segmentation algorithm that produces segmentation results and a region-based hierarchical tree concurrently in an efficient and effective way. In addition, introducing the region-based hierarchical tree can preserve the relations among segmented regions, which also ease the over-segmentation issue in the subsequent object matching process. Further, the proposed sub-tree comparison approach provides an efficient way of performing object matching and multi-object retrieval. Moreover, we maximize the usage of the user's feedback in query refinement and avoid the expensive proper sub-tree comparison in the feedback process. By means of the seamless integration of the user's relevance feedback into the proposed MOIR system, it allows automatic discovery of the object(s) of the user's interest and improves the retrieval accuracy through feedback-retrieval loops.

REFERENCES

[1] B. Luo, X. Wang, and X. Tang, "A world wide web based image search engine using text and image content features," Proc. of the 2003 SPIE-IS&T Electronic Imaging, Internet Imaging IV, 2003, pp. 123-130.

[2] T. Kato, "Database architecture for content-based image retrieval," Proc. of SPIE Conference on Image Storage and Retrieval Systems, 1992, vol. 1662, pp.112-130.

[3] Y. Chen, J. Li, and J. Z. Wang, "Machine learning and statistical modeling approaches to image retrieval," Norwell, Massachusetts: Kluwer Academic Publishers, 2004.

[4] Y. Rui, T.S. Huang, and S.-F. Chang, "Image retrieval: past, present and future," Journal of Visual Communication and Image Representation, 1997, 10, pp. 1-23.

[5] H. Shao, Y. Wu, W. Cui, and J. Zhang, "Image retrieval based on MPEG-7 dominant color descriptor," Proc. of the 9th International Conference for Young Computer Scientists (Zhang Jia Jie, Hunan, China, November 18 – 21). ICYCS'08 , 2008, pp. 753-757.

[6] J. Li, J. Z. Wang, and G. Wiederhold, "IRM: Integrated region matching for image retrieval," Proc. of the 2000 ACM international conference on Multimedia, 2000, pp. 147-156.

[7] R. Ji, H. Yao, and D. Liang, "DRM: Dynamic region matching for image retrieval using probabilistic fuzzy matching and boosting feature selection," Signal, Image and Video Processing, 2008, 2, pp. 59-71.

[8] C. Zhang, L. Zhou, W. Wan, J. Birch, and W.-B. Chen, "An image clustering and feedback-based retrieval framework," International J. of Multimedia Data Engineering and Management, 2010, 1, 1, pp. 55-74.

[9] P. Li, J. Guo, B. Song, and X. Xiao, "A multilevel hierarchical image segmentation method for urban impervious surface mapping using very high resolution imagery," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 2011, 4, 1, pp. 103-116.

[10] F. S. Al-Qunaieer, H. R. Tizhoosh, and S. Rahnamayan, "Multi-resolution level set image segmentation using wavelets," Proc. of the IEEE International Conference on Image Processing (ICIP'11), 2011, pp. 269-272.

[11] V. Vilaplana and F. Marques, "On building a hierarchical region-based representation for generic image analysis," Proc. of the IEEE International Conference on Image Processing (ICIP'07), 2007, 4, pp. 325-328.

[12] B. Sumengen and B. S. Manjunath, "Multi-scale edge detection and image segmentation," Proc. of the 2005 European Signal Processing Conference (EUSIPCO'05), 2005.

[13] D. Prewer and L. Kitchen, "Soft image segmentation by weighted linked pyramid," Pattern Recognition Letters, 2001, 22, pp. 123-132.

[14] Y. Xu, P. Duygulu, E. Saber, A. M. Tekalp, and F. T. Yarman-Vural, "Object based image retrieval based on multi-level segmentation," Proc. of the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP'00), IEEE Computer Society, 2000, 4, pp. 2019-2022.

[15] P. J. Burt, T. H. Hong, and A. Rosenfeld, "Segmentation and estimation of image region properties through cooperative hierarchical computation," IEEE Transactions on Systems, Man, and Cybernetics, 11, 12, 1981, pp. 802-809.

[16] N. Ahuja and S. Todorovic, "Connected segmentation tree – a joint representation of region layout and hierarchy," Proc. of the Computer Vision and Pattern Recognition (CVPR'08), 2008, pp.1-8.

[17] P. Arbelaez, M. Maire, C. Fowlkes, and J. Malik, "From contours to regions: an empirical evaluation," Proc. of the Computer Vision and Pattern Recognition (CVPR'09), 2009, pp. 2294-2301.

[18] A. Haar, "Zur Theorie der orthogonalen Funktionensysteme," Mathematische Annalen. 69, 3, 1910, pp. 331-371.

[19] H. Shao, Y. Wu, W. Cui, and J. Zhang, "Image retrieval based on MPEG-7 dominant color descriptor," Proc. of the 9th International Conference for Young Computer Scientists (ICYCS'08), 2008, pp. 753-757.

[20] R. Adams and L. Bischof, "Seeded region growing," IEEE Trans. on Pattern Analysis and Machine Intelligence, 1994, 16, 6, pp. 641-647.

[21] J. Z. Wang, J. Li, and G. Wiederhold, "SIMPLIcity: Semantic-Sensitive Integrated Matching for Picture Libraries," IEEE Trans. on Pattern Analysis and Machine Intelligence, 2001, 23, 9, pp. 947-963.

[22] C. Carson, S. Belongie, H. Greenspan, and J. Malik, "Blobworld: image segmentation using expectation-maximization and its application to image querying," IEEE Trans. on Pattern Analysis and Machine Intelligence, 2002, 24, 8, pp. 1024-1038.

[23] B. Schölkopf, J. C. Platt, J. C. Shawe-Taylor, A. J. Smola, and R. C. Williamson, "Estimating the support of a high-dimensional distribution," Microsoft Research Corporation Technical Report MSR-TR-99-87, 1999.

[24] T. Ohashi, Z. Aghbari, and A. Makinouchi, "Hill-climbing algorithm for efficient color-based image segmentation," Proc. of the IASTED International Conference on Signal Processing, Pattern Recognition, and Applications (SPPRA'03), 2003.

[25] R. Achanta, F. Estrada, P. Wils, and S. Süsstrunk, "Salient region detection and segmentation," Proc. of the International Conference on Computer Vision Systems (ICVS'08), 2008.