An Innovative Multiple-Object Image Retrieval Framework Using Hierarchical Region Tree

Wei-Bang Chen, Department of Mathematics and Computer Science, Virginia State University, Petersburg, VA, USA

Chengcui Zhang, Department of Computer and Information Sciences, University of Alabama at Birmingham, Birmingham, AL, USA

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. This paper presents an innovative multiple-object retrieval framework named Multiple-Object Image Retrieval (MOIR) on the basis of hierarchical image representation. This framework 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 designed for multiple-object retrieval with a reasonably low time complexity. The experimental results demonstrate the efficacy and efficiency of the proposed approach.

Keywords: *Content-Based Image Retrieval, Hierarchical Region-Tree, Image Segmentation, Multi-Object Retrieval, Multi-Resolution Image Segmentation*

1. INTRODUCTION

The evolution of digital technology promotes information storage migrating from analogue to digital form, and results in convenient information sharing and distribution (Li et al., 2000). Since 1980, the digital revolution has driven the explosion of digital devices on the market, which makes digital imaging emerge from its infancy in the past decade. 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).

DOI: 10.4018/jmdem.2013070101

In most conventional text-based image search systems such as Flickr, 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 (Luo et al., 2003). 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 (Luo et al., 2003), 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 (Kato, 1992). 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 (Chen et al., 2004). 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 (Rui et al., 1997).

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 (Samadani et al., 1993; Pentland et al., 1994; Kelly and Cannon, 1995; Stone and Li, 1996; Wong and Po, 2004). In general, full image search is relatively simple and efficient, but less human-centered. The reason is that humans find images based on the high level concepts, such as objects or scenes; however, global visual features used in the full image search cannot capture the properties of those high level concepts.

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. In order to perform object-based search, it is essential to extract objects embedded in images. This extraction process is called image segmentation which splits images into meaningful regions, each of which represents a constituent object.

Image segmentation is known to be one of the most challenging issues in the field of image processing. Many efforts have been made to improve the segmentation accuracy. Most segmentation algorithms distinguish image segments on the basis of color (Lucchese and Mitra, 2001; Rahimizadeh et al., 2009), texture (Xie and Mirmehdi, 2007), and/or edge (Yu and Clausi, 2008; Arbeláez et al., 2009). Only a few segmentation algorithms combine multiple visual features (Deng and Manjunath, 2001; Carson et al., 2002; Kato and Pong, 2006; Kumar et al., 2008). In addition, image segmentation is known to be a time- or storage-consuming process (Sadek et al., 2009).

Another challenging issue in image segmentation is over-segmentation 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. Over-segmented regions indicate that an object is divided into two or more smaller segments. On the other hand, under-segmented regions indicate that two or more objects are merged into a larger segment. 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 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. (2000) 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. Another approach, dynamic region matching (DRM) was also introduced in 2008 by Ji et al. to address the same issue (Ji et al., 2008). In 2010, Zhang et al. proposed a feedback-based image clustering and retrieval framework (FIRM), which has demonstrated that combining integrated region matching (IRM) scheme with users' relevance feedback (RF) in a multi-object based image retrieval framework makes the automatic discovery of user desired objects possible. However, both IRM and DRM based 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 (Xu et al., 2000; Prewer and Kitchen, 2001; Sumengen and Manjunath, 2005; Vilaplana and Marques, 2007; Al-Qunaieer et al., 2011; Li et al., 2011). 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 framework 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 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.

2. PROPOSED METHOD

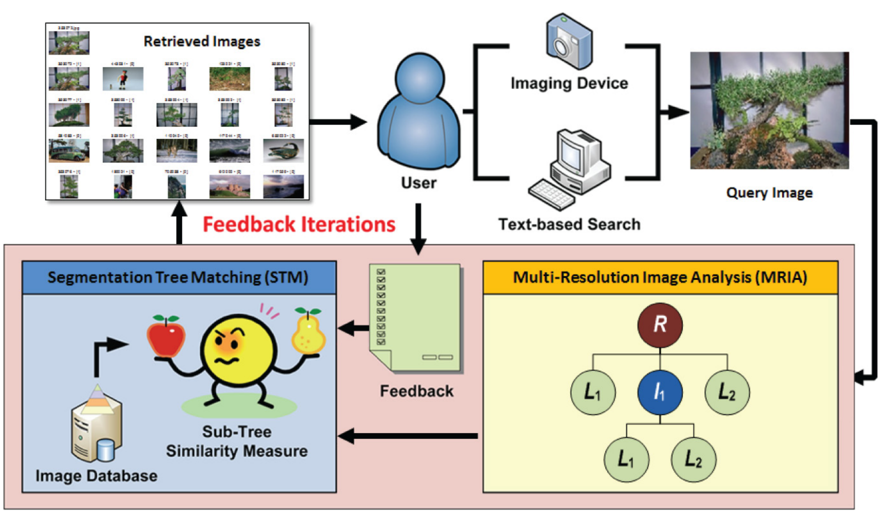
2.1. Framework Overview

The proposed multiple-object image retrieval (MOIR) framework adopts the query by example (QBE) technique, and thus, it starts database search with the submission of a query sample image. Next, the MOIR framework performs the proposed multi-resolution image analysis (MRIA) on the query sample image in order to perform image segmentation and build a hierarchical region tree in a concurrent fashion. In the next step, the 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. Subsequently, target images are ranked in the descending order of their similarities to the query image. Then, the top 20 images in the ranked list are retrieved and displayed to the user for reviewing and leaving feedback to the retrieval system. In order to refine the retrieval results on the basis of users' feedback, the relevance feedback (RF) technique is also adopted in the framework. The high level architecture of the proposed MOIR framework is illustrated in Figure 1.

2.2. Hierarchical Image Representation

To perform object-based image retrieval, it is essential to extract meaningful objects contained in images. As aforementioned, the main challenging issue in object-based image retrieval is how to alleviate the problem of inaccurate segmentation. An inaccurately segmented image suffers from both over- and under-segmentation problems. Both over-segmentation and under-segmentation have negative impacts on the retrieval accuracy in object-based image retrieval. Researchers introduce hierarchical image representation to preserve the relationship between and among segments (Burt et al., 1981; Ahuja, 2008; Arbeláez et al., 2009) in order to reduce the negative impact of inaccurate segmentation in object-based image retrieval. The

Figure 1. The high level view of the multiple-object image retrieval (MOIR) framework



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.

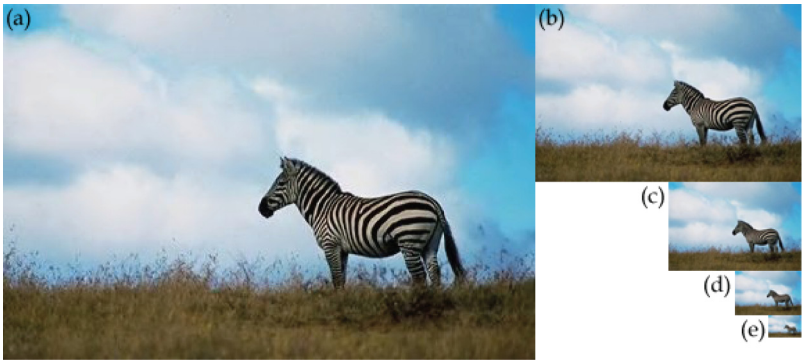
In this paper, we proposed a multi-resolution image analysis (MRIA) algorithm that performs hierarchical image segmentation with the above desired characteristics. The proposed multi-resolution image analysis (MRIA) algorithm is inspired by the human visual system. Imagining you are standing on an open field and a red sports car is moving toward you from 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 multi-resolution image analysis (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.

Figure 2 exemplifies a series of multi-resolution images where (a) is the original image, (b) – (e) is 1/2, 1/4, 1/8, and 1/16 down-sampled image of the original image in each dimension, respectively. In Figure 2 (a), the high-frequency signals such as the black and white stripes on the zebra are distinct. After a

Figure 2. A series of multi-resolution images



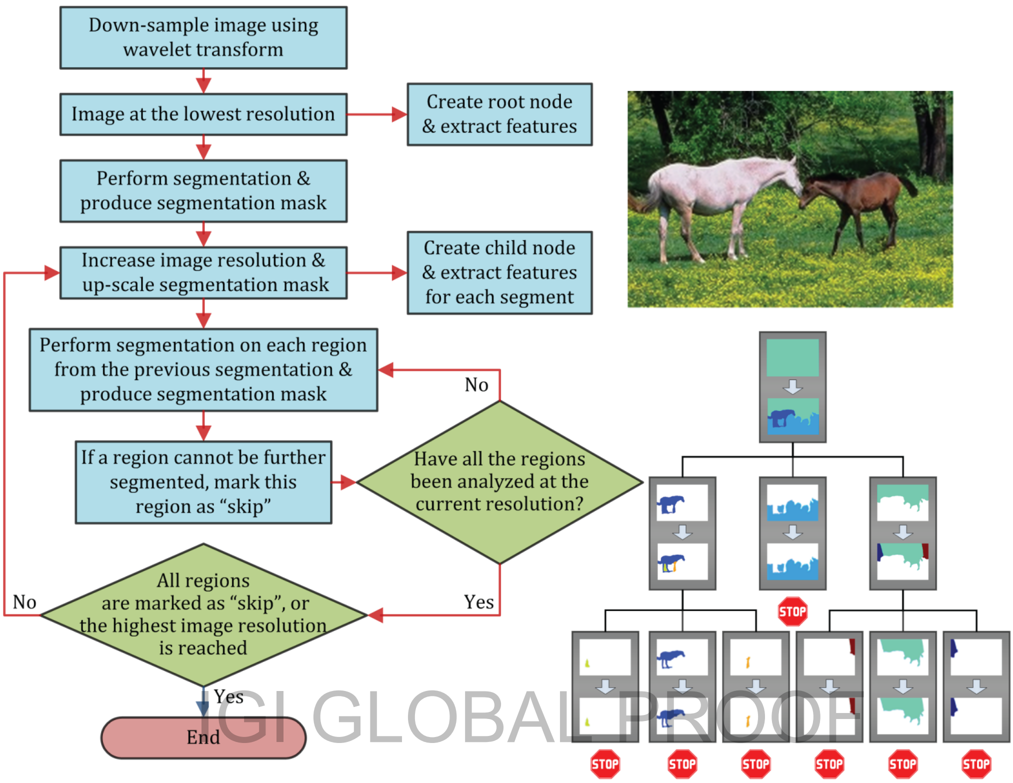
series of down-samplings, the black and white stripes on the zebra become blurred in (c), and totally vanish in (d) & (e). This indicates that the high-frequency signals are removed from the image in the process of down-sampling.

The flowchart of the proposed MRIA algorithm is depicted in Figure 3. As aforementioned, one of our goals 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 (Haar, 1910) 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 (Shao et al., 2008). The motivation is that dominant color descriptor is efficient and effective in describing the entire or a portion of an image with representative color distributions.

By considering the entire image as one region, the first level image segmentation (region growing algorithm (Adams et al., 1994; Shih et al., 2005) is used in this paper) is performed on the region, producing a segmentation mask. Then, we increase the image resolution and upscale 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

Figure 3. The multi-resolution image analysis framework for hierarchical image representation

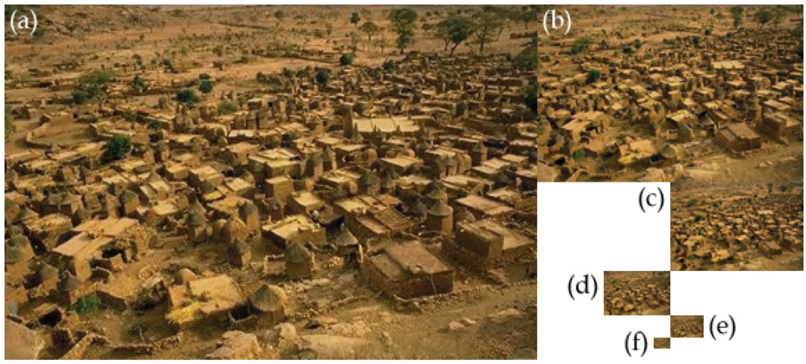


will likely to be further segmented into smaller segments at a higher image resolution, and a new segmentation mask is produced for that region. This process will continue until either of the following criteria is satisfied. The two stopping criteria are: (1) no region can be further segmented, or (2) the highest image resolution is reached.

However, about 4 percent of images in our dataset cannot be properly processed using the aforementioned multi-resolution image analysis (MRIA) algorithm. The reason is that those images usually contain highly similar foreground and background such that the multi-resolution image analysis stopped at the lowest image resolution due to the fact that no region can be further segmented. For instance, Figure 4 depicts a town with mud houses and a few trees. The original image (size: 384×256) is displayed in (a), and a series of images at different resolutions is demonstrated in (b) to (f). In the multi-resolution images, (b)-(e) are intermediate resolution images, and (f) is an image at the lowest solution (12×8). We can observe from these images that image details below certain resolution are almost completely removed and the differences between foreground and background are invisible, in particular (e) and (f). This example suggests that in some images, objects such as the foreground and the background cannot be well differentiated at low image resolutions, and thus, it is not reasonable to immediately stop the multi-resolution image analysis process if segmentation at the current resolution exposes no object at all.

In order to cope with this issue, we slightly modify the process of multi-resolution image analysis by relaxing the stopping criteria for processing images. Specifically, the modified algorithm will not apply the criterion of "no

Figure 4. A sample image that contains highly similar foreground and background



region can be further segmented" until the total number of segmented regions in the image is ≥ 2. This provides a chance for objects invisible to a low resolution analysis to be extracted at a higher resolution where more details become available.

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 manner, the computational cost can be dramatically reduced.

Although the image hierarchy is used 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.

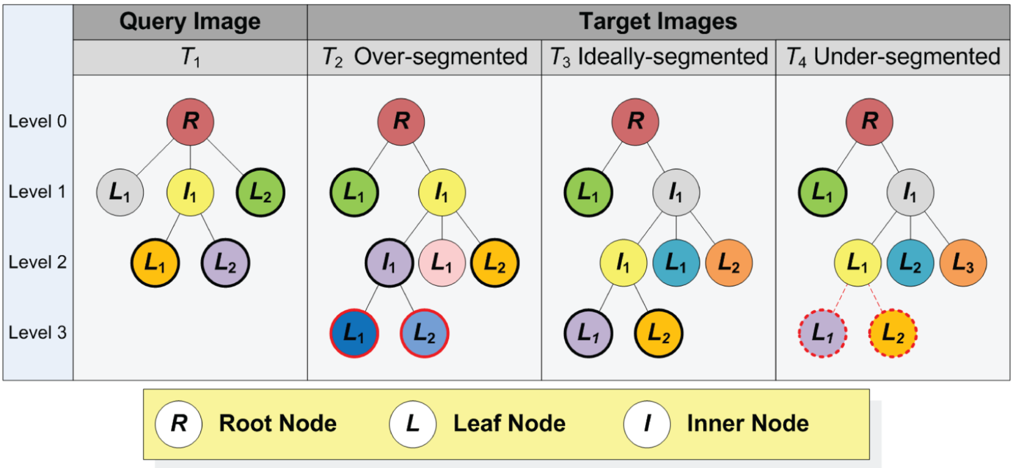
2.3. Hierarchical Region Tree Matching

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 5 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 5 demonstrates four hierarchical region trees T_1, T_2, T_3 and T_4 which model the

Figure 5. Four examples of hierarchical region trees



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 5, 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 (Wang et al., 1999; Wang et al., 2001). 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 subtree 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 5 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 5, 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 do not 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).

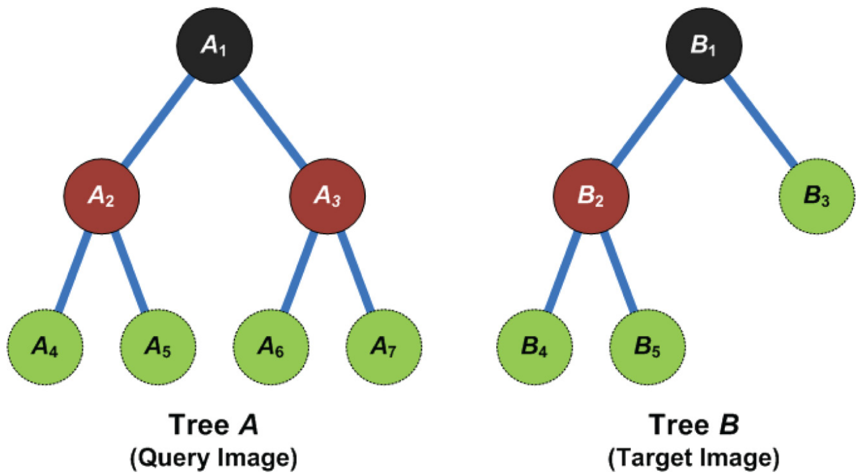
From the above examples, three types of comparison can be concluded, including leaf to leaf (L-L) comparison, leaf to sub-tree or sub-tree to leaf (L-P/P-L) comparison, and sub-tree to sub-tree (P-P) comparison. The above three types of comparisons are actually performed through measuring the similarity 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 (Carson et al. 2002). 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 subtree similarity based on previously calculated similarity values during subsub-tree comparison, similar to the idea of dynamic programming. We use the following example (as shown in Figure 6) to explain the proposed segmentation tree matching algorithm.

Figure 6 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 reversed 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

Figure 6. Matching two hierarchical region trees



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.

2.4. 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 propose to adopt one-class support vector machine (SVM) (Schölkopf et al., 1999) 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.

In summary, by addressing and attempting to solve the above challenging issues, we expect that the development of the multi-resolution image analysis (MRIA) will provide us with an efficient tool to simultaneously produce image segmentation results and hierarchical region-tree representations, which are typically obtained through two separate processes with existing approaches. The hierarchical image representation is expected to alleviate the object matching problem due to the negative impact of over-segmentation. Further, with such a hierarchical representation, the relevance of a target image to the query image, in terms of their object similarity, is thus measured according to their proper sub-tree similarity. In the proposed framework, we also design and develop an efficient strategy to compare proper sub-trees. An innovative relevance feedback scheme is also proposed to bridge the semantic gap, providing the capability for the system to learn the object(s) of the user's interest with a very small training set, which maximizes the usage of the user's feedback in query refinement and avoids the expensive proper sub-tree comparison. By means of the seamless integration of users' relevance feedback (RF) into the proposed multiple-object image retrieval (MOIR) system, it allows automatic discovery of the objects of the user's interest and improves the retrieval accuracy through feedback-retrieval loops.

3. EXPERIMENTAL RESULTS

3.1. Dataset Description

The experiments in this paper are performed on a dataset containing 10,000 nature scene images originated from 100 theme categories in Corel Photo CDs. It is worth mentioning that we use Corel image dataset in a different way from conventional methods. Unlike the traditional way where Corel theme category labels are used as ground-truth, we procure our own ground-truth for evaluation. Specifically, we define 54 objects of interests and manually annotate images containing the objects of interests. Many of these objects (e.g., blue sky, red car, snow, fence, lighthouse, horse and roadway) occur in multiple Corel theme categories. Our ground-truth labels are these manually annotated objects instead of Corel theme category labels. The number of images for each object of interest used in the experiments is demonstrated in Table 1.

3.2. Complexity Analysis

3.2.1. Multi-Resolution Image Analysis (MRIA) Algorithm

For an original image of size $m \times n$, assume that the number of down-sampling (decimation by a factor of 2) levels is k. The level-k represents the down-sampled image at the second highest resolution (the original image has the highest resolution). The size of the down-sampled image at the lowest resolution (level-1) is $a \times b$ where $a = \frac{m}{2^k}$ and $b = \frac{n}{2^k}$.

For the proposed multi-resolution image analysis algorithm, the worst case is when we need to segment the entire image for all of the down-sampled images and the original image. In the worst case $\left(k \to \infty\right)$, the total number of pixels needed for processing is:

Table 1. Object of interest statistics

Object	# of Image	Object	# of Image	Object	# of Image
blue sky	1810	lighthouse	110	roadway	319
bonsai	100	martial arts	100	shoji	18
bullet	23	penguin	66	snow	584
dinosaur	101	pyramid	24	tiger	102
fence	243	red airplane	28	white rabbit	16
gun	76	red bus	53	yellow bus	23
horse	244	red car	109	yellow car	34

$$a \times b + 2^1 a \times 2^1 b$$
$$+2^2 a \times 2^2 b + \cdots + 2^k a \times 2^k b$$

where the first term $a \times b$ indicates the number of pixels at the lowest resolution (level-1), the second term $2^1 a \times 2^1 b$ indicates the number of pixels at the second lowest resolution (level-2), and the last term $2^k a \times 2^k b$ indicates the number of pixels at the highest resolution, i.e., the original image.

Introducing $a = \frac{m}{2^k}$ and $b = \frac{n}{2^k}$ into the above equation, we can obtain the following equation:

$$\frac{m}{2^k} \times \frac{n}{2^k} + \frac{m}{2^{(k-1)}} \times \frac{n}{2^{(k-1)}}$$
$$+\frac{m}{2^{(k-2)}} \times \frac{n}{2^{(k-2)}} + \cdots + \frac{m}{2^0} \times \frac{n}{2^0} \approx \mathrm{O}\left(1.3mn\right)$$

The above equation indicates that the total number of pixels needed for processing in the worst case is $\mathrm{O}\left(1.3mn\right)$.

Based on our experiment, the maximum value of *k* is 6, and the proposed multi-resolution image analysis (MRIA) algorithm stops by level 3 for more than 57% images and by level 4 for more than 98% images. This indicates that the proposed multi-resolution image analysis (MRIA) algorithm is very efficient in performing image segmentation and tree construction since in most cases the proposed multi-resolution image analysis (MRIA) algorithm only needs to process about 2% (level 3) to 8% (level 4) of the original image size.

3.2.2. Hierarchical Region Tree Comparison Algorithm

As aforementioned, there are three types of tree node comparisons, including: leaf to leaf comparison (L-L), sub-tree to leaf comparison (P-L), and sub-tree to sub-tree comparison (P-P). The proposed hierarchical region tree comparison algorithm performs comparison not only on L-L but also on P-L and P-P whose complexity is determined by the number of inner nodes. Given the same number of leaf nodes, a binary tree has the most inner nodes, thus the worst case is the comparison of two binary trees.

Assume the numbers of leaf nodes in binary trees *A* and *B* are *m* and *n*, respectively. In a binary tree, the number of inner nodes is the number of leaf nodes minus 1, and thus, the numbers of inner nodes in *A* and *B* are *m* – 1 and *n* – 1, respectively.

The time complexity of each type of tree node comparisons is calculated as follows:

L-L: $m \times n$

P-L: $\left(m-1\right) \times n$ /* without loss of generality, assume we compare all the sub-trees in A with all the leaf nodes in B */

P-P: $\left(m-1\right) \times \left(n-1\right)$

Therefore, the overall complexity is the sum of the three types of tree node comparisons, which can be obtained from the following calculation:

$$m \times n + \left(m-1\right) \times n + \left(m-1\right) \times \left(n-1\right)$$
$$= mn + mn - n + mn - m - n + 1$$
$$= 3mn - 2n - m + 1$$
$$\approx \mathrm{O}\left(mn\right)$$

From the above equation, the worst case overall time complexity for the proposed hierarchical region tree comparison is $\mathrm{O}\left(mn\right)$.

Based on our experiment, the average number of inner nodes in our dataset is about 3.2. Further, while an inner node has a branching factor of at least 2 in our case, many inner nodes have more than two children, suggesting that the actual total number of P-L and P-P comparisons is even lower.

3.3. 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.

3.3.1. MRIA Efficiency Analysis

In this study, we propose a novel measure named segmentation efficiency as defined below for the following experiment. 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:

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

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 multi-resolution image analysis (MRIA) is 98.26%. This indicates that our approach is very efficient in segmenting an image and constructing the hierarchical image representation in one run.

3.3.2. MRIA Efficacy Analysis

We introduce a subjective quality assessment experiment to evaluate the efficacy of the proposed multi-resolution image analysis (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) (Ohashi et al., 2003; Achanta et al., 2008). To ensure the integrity of subjective evaluation, 9 evaluators perform a blind review through a web interface as shown in Figure 7. The evaluators vote the best segmented image from the two displayed segmented images produced by MRIA and HCK algorithms, respectively. The evaluation system also provides a neutral option, if both segmented images are comparable. The results of the assessment are presented in Table 2.

In Table 2, the numbers '8' and '10' represent the different number of bins used in the HCK algorithm. Table 2 demonstrates that the image segmentation quality of the proposed MRIA algorithm significantly outperforms the HCK algorithm.

Figure 7. The blind review web interface for efficacy evaluation

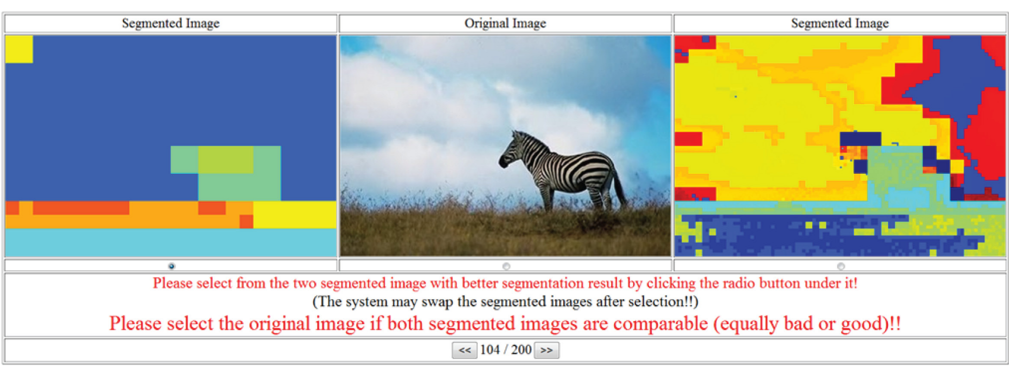


Table 2. Subjective segmentation quality assessment

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

3.4. Evaluation Metrics for Multiple Object Image Retrieval

Two commonly used standard evaluation metrics, Average Precision (AveP) and Mean Average Precision (MAP), are used in the following experiments in order to fairly compare the proposed multiple object image retrieval (MOIR) framework to other existing frameworks (TRECVID, 2009). 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.

The definition of the average precision (AveP) measure for a query k is formalized in the following equation:

$$AveP\left(k\right) = \frac{\sum_{s=1}^{m}\left(P\left(s\right)\times rel\left(s\right)\right)}{number\, of\, retrieved\, relevant\, documents}$$

where s is the rank in the sequence of the retrieved list, and m is the number of retrieved images. $P(s)$ indicates the precision at cut-off s in the list, and $rel(s)$ represents an indicator function equaling 1 if the item (image) at rank s is relevant, zero otherwise.

The mean average precision (MAP) is the arithmetic mean of the average precision values obtained from a set of queries, which is defined in the equation below:

$$MAP = \frac{\sum_{k=1}^{N} AveP\left(k\right)}{N}$$

where N is the number of queries.

Using these measures provides an objective and comprehensive view when comparing the performance of the proposed framework to other existing approaches.

3.5. MOIR Framework Assessment

Based on the MAP measure, we compare the proposed MOIR framework to three state-of-the-art object-based image retrieval frameworks, including integrated region matching (IRM) (Li et al., 2000), feedback-based image clustering and retrieval framework (FIRM) (Zhang et al., 2010), and dynamic region matching (DRM) (Ji et al., 2008). 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.

3.5.1. Comparison of Image Matching Algorithms

In Section 3.3, we have demonstrated the efficacy and efficiency of the proposed MRIA algorithm for image segmentation and hierarchical region tree construction

However, without an efficient and effective hierarchical region tree comparison algorithm, using hierarchical image representation alone cannot relieve the problem of imperfect image segmentation. In this research, we develop an efficient hierarchical region tree comparison algorithm, namely segmentation tree matching (STM), to work together with hierarchical image representation in order to ease the problem of over-segmentation in object-based image retrieval.

To assess the effectiveness of the proposed segmentation tree matching (STM) algorithm, we compare the proposed segmentation tree matching (STM) algorithm to several state-of-the-art approaches. These state-of-the-art approaches we use for comparison include the earth mover's distance (EMD) (Pele & Werman, 2009), the integrated region matching (IRM, the core algorithm of SIMPLIcity, Stanford Univ.) (Li, et al., 2000), feedback-based image clustering and retrieval framework (FIRM) (Zhang, et al., 2010), and dynamic region matching (DRM) (Ji, et al., 2008).

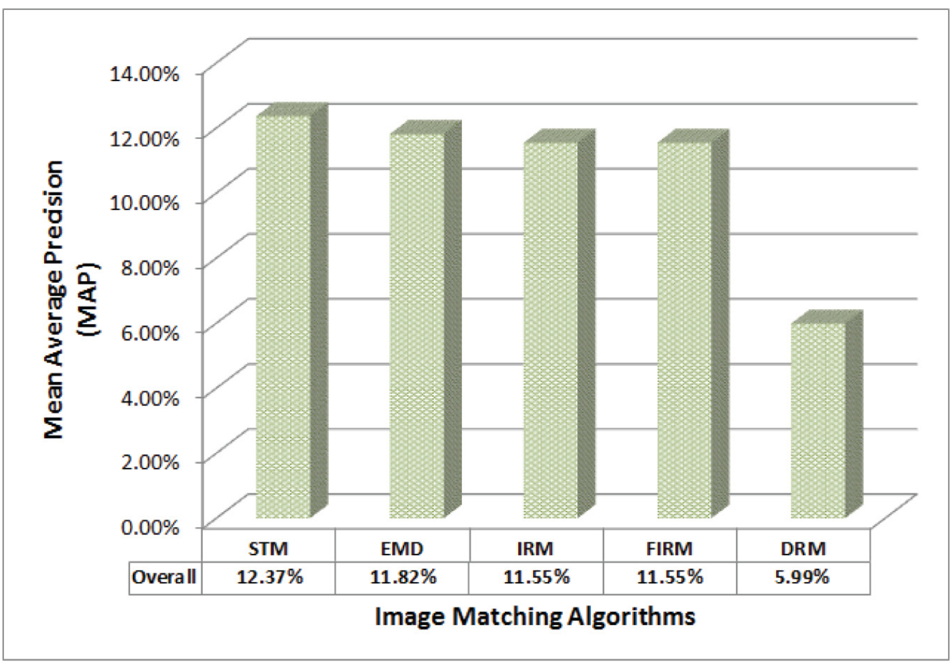
The assessment also adopts the MAP value as evaluation metric and compares the performance of STM to that of the state-of-the-art algorithms. As demonstrated in Figure 8, STM has the highest overall MAP value (12.37%), followed by EMD (11.82%), IRM (11.55%), and FIRM (11.55%). DRM has a significantly lower overall average MAP value (5.99%) than the others. Among the results produced by five different algorithms, STM has the best performance.

3.5.2. Single-Object Image Retrieval

To learn users' query intention, it would be much easier if an image retrieval system requests a user to explicitly specify the object(s) of his/her interest. However, it would be not only troublesome but also impractical for a user to put so many efforts in the user interaction cycle. For example, some existing systems request users to select the regions that correspond to desired objects, but it would be very difficult and confusing for users to complete this requirement when the image is over-segmented or under-segmented., let alone asking the user to select multiple objects from such an image.

In the proposed multiple-object image retrieval (MOIR) framework, our goal is to minimize the amount of user effort, from which essential information for discovery of user-desired objects can be procured by the proposed approach. For this purpose, we introduce the relevance feedback (RF) technique,

Figure 8. Comparison of image matching algorithms



a supervised machine learning mechanism, to analyze users' feedback.

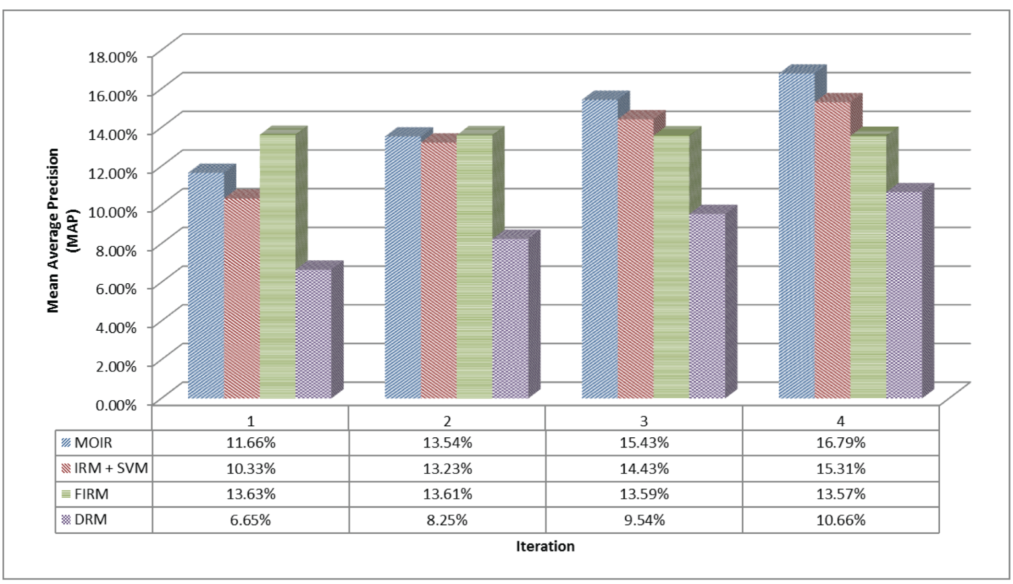
In object-based image retrieval, retrieving a single object is the most fundamental function. In this experiment, we demonstrate the effectiveness of the proposed MOIR framework in single object retrieval based on 770 query images from 13 categories including dinosaur (100/100), red bus (53/53), pyramid (24/24), white rabbit (16/16), bullet (23/23), yellow car (34/34), yellow bus (23/23), bonsai (100/100), tiger (102/102), penguin (66/66), shoji (18/18), lighthouse (110/110), and martial arts (100/100), where the first number in the parentheses represents the number of query images in each category and the second number in the parentheses represents the total number of images containing the query object in the dataset. The retrieval performance of each framework in terms of MAP is shown in Figure 9.

Figure 9 summarizes the retrieval performance of each framework in terms of MAP. The blue bar represents the proposed multiple-object image retrieval (MOIR) framework; the red bar represents the IRM with SVM; the green bar represents the FIRM; and the purple bar represents the DRM. From Figure 9, it can be observed that after 4 feedback iterations, the MAP value of the proposed MOIR framework reaches 16.79%, which is 1.48%, 3.22%, and 6.13% higher than that of the IRM+SVM, FIRM, and DRM, respectively. The MOIR frame-work significantly outperforms other frameworks since we conduct a census-based evaluation. In a census-based evaluation, we had eliminated the bias and error caused by random sampling, which means any increase is statistically significant. Further, the MAP value of MOIR steadily increases through the feedback iterations, which indicates the robustness and effectiveness of the relevance feedback.

3.5.3. Multiple-Object Image Retrieval

Similar to the single object image retrieval, in this experiment we demonstrate the effectiveness of the MOIR framework in multiple-object retrieval based on 398 query images from 12 different query object combinations, including: bonsai + shoji (18/18), blue sky + red bus (34/34), pyramid + blue sky (21/21), white rabbit + snow (11/11), gun + bullet (19/19), red

Figure 9. Single-object retrieval results (770 queries)



airplane + blue sky (18/18), red car + roadway (45/45), yellow bus + roadway (19/19), yellow car + roadway (16/16), lighthouse + blue sky (70/70), fence + blue sky (47/47), and fence + horse (80/80), where the first number in the parentheses represents the number of query images in that combination and the second number in the parentheses represents the total number of images containing all the query objects in the dataset. In this experiment, we again adopt the MAP value as the evaluation metric and compare the performance of the MOIR framework (STM+SVM), to that of the state-of-the-art approaches. The performance for each framework is shown in Figure 10.

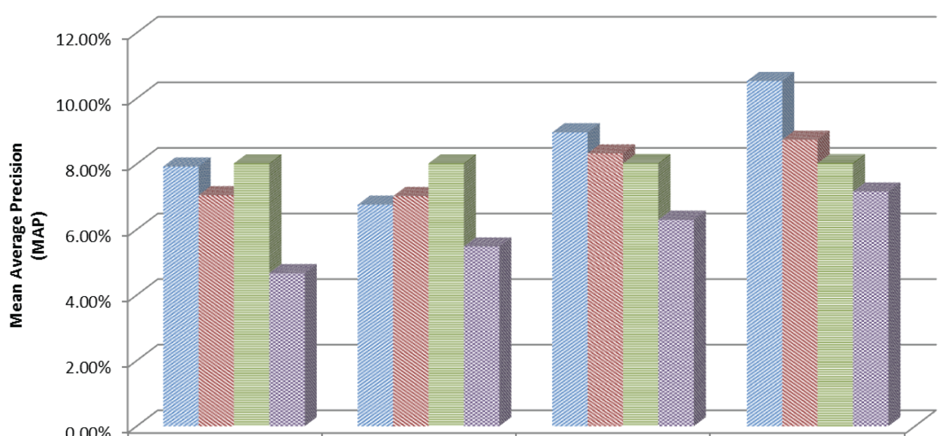
The performance for each framework is summarized in Figure 10 which shows that after four feedback iterations, the proposed MOIR framework significantly outperforms IRM+SVM, FIRM, and DRM by 1.79%, 2.51%, and 3.37%, respectively. Similarly, the MAP value of our algorithm also increases through the feedback iterations, again indicating the effectiveness and robustness of the relevance feedback. In addition, Figures 11 (a) and (b) exemplify the top 20 retrieved images after the 1st feedback and the 4th feedback iterations for a sample query image and their corresponding AveP values, respectively. The query image is displayed on the top left corner.

4. CONCLUSION

In this paper, we introduce an innovative human-centered multiple-object image retrieval framework (MOIR) based on hierarchical image representation. The novelties of the proposed framework are manifold.

First, this framework seamlessly integrates a multi-resolution hierarchical segmentation algorithm (MRIA) that produces segmentation results and a region-based hierarchical tree concurrently in an efficient and effective manner. The proposed MRIA is different from existing hierarchical image segmentation. Existing methods need to perform full segmentation at each resolution, while MRIA segments an

Figure 10. Multiple-object retrieval results (398 queries)

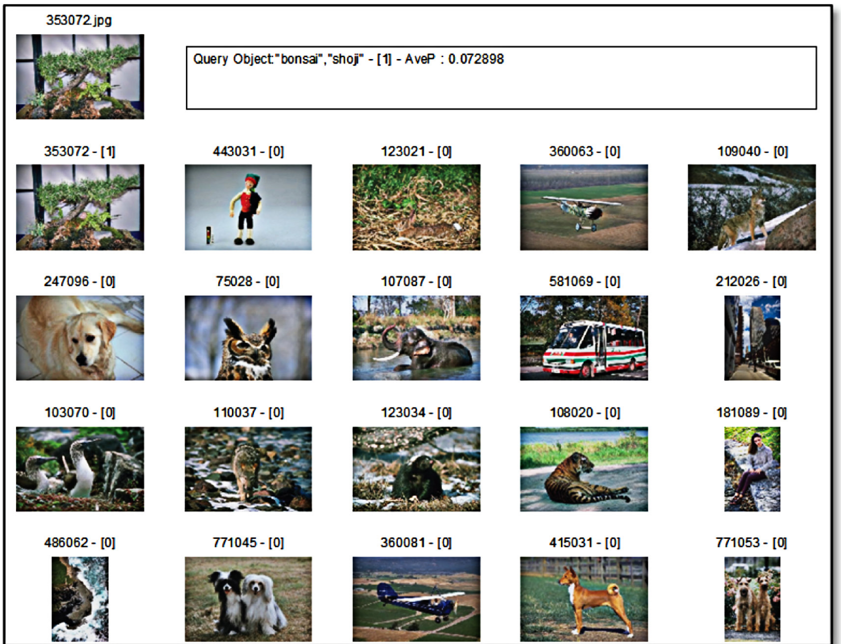


	1	2	3	4
MOIR	7.92%	6.76%	8.96%	10.53%
IRM + SVM	7.05%	7.02%	8.32%	8.74%
FIRM	8.02%	8.02%	8.02%	8.02%
DRM	4.67%	5.48%	6.30%	7.16%

Iteration

Figure 11. MOIR retrieval results: (a) 1st feedback; (b) 4th feedback

(a)



(b)

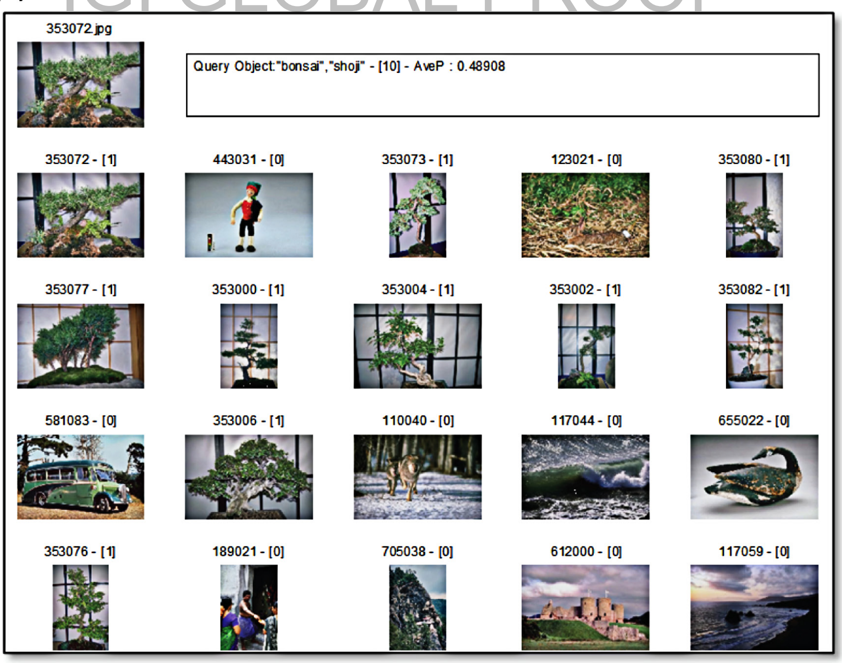


image starting at a very low resolution, and proceeds to segmentation at a higher resolution only when needed. Therefore, MRIA is more efficient than existing methods.

Second, 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.

Third, the proposed segmentation tree matching (STM) algorithm 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 (RF) into the proposed multiple-object image retrieval (MOIR) system, it allows automatic discovery of the object(s) of the user's interest and improves the retrieval accuracy through feedback-retrieval loops.

In summary, to our best knowledge we are the first to explore the combined problem of hierarchical image segmentation, representation, and customized object-based image retrieval and to have made non-trivial progress.

REFERENCES

Achanta, R., Estrada, F., Wils, P., & Süsstrunk, S. (2008). Salient region detection and segmentation. In *Proceedings of the International Conference on Computer Vision Systems (ICVS 2008)*.

Adams, R., & Bischof, L. (1994). Seeded region growing. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *16*(6), 641–647. doi:10.1109/34.295913

Ahuja, N., & Todorovic, S. (2008, June 23–28). Connected segmentation tree – a joint representation of region layout and hierarchy. In *Proceedings of the Computer Vision and Pattern Recognition (CVPR 2008)*, Anchorage, AK (pp. 1-8).

Al-Qunaieer, F. S., Tizhoosh, H. R., & Rahnamayan, S. (2011). Multi-resolution level set image segmenation using wavelets. In *Proceedings of the IEEE International Conference on Image Processing (ICIP 2011)*, Brussels, Belguim (pp. 269-272).

Arbeláez, P., Maire, M., Fowlkes, C., & Malik, J. (2009). From contours to regions: an empirical evaluation. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (pp. 2294-2301).

Burt, P. J., Hong, T. H., & Rosenfeld, A. (1981). Segmentation and estimation of image region properties through cooperative hierarchical computation. *IEEE Transactions on Systems, Man, and Cybernetics*, *11*(12), 802–809. doi:10.1109/TSMC.1981.4308619

Carson, C., Belongie, S., Greenspan, H., & Malik, J. (2002). Blobworld: Image segmentation using expectation-maximization and its application to image querying. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *24*(8), 1024–1038. doi:10.1109/TPAMI.2002.1023800

Chen, Y., Li, J., & Wang, J. Z. (2004). *Machine learning and statistical modeling approaches to image retrieval*. Norwell, MA: Kluwer Academic Publishers.

Deng, Y., & Manjunath, B. S. (2001). Unsupervised segmentation of color-texture regions in images and video. [PAMI]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *23*(8), 800–810. doi:10.1109/34.946985

Haar, A. (1910, September). Zur Theorie der orthogonalen Funktionensysteme. *Mathematische Annalen*, *69*(3), 331–371. doi:10.1007/BF01456326

Ji, R., Yao, H., & Liang, D. (2008). DRM: Dynamic region matching for image retrieval using probabilistic fuzzy matching and boosting feature selection. *Signal*. *Image and Video Processing*, *2*, 59–71. doi:10.1007/s11760-007-0037-0

Kato, T. (1992). Database architecture for content-based image retrieval. In *Proceedings of the SPIE Conference on Image Storage and Retrieval Systems, 1662*, 112–130.

Kato, Z., & Pong, T.-C. (2006). A Markov random field image segmentation model for color textured images. *Image and Vision Computing*, *24*, 1103–1114. doi:10.1016/j.imavis.2006.03.005

Kelly, P. M., & Cannon, M. (1995). Query by image example: the comparison algorithm for navigating digital image databases (CANDID) approach. In *Proceedings of the SPIE Conference on Storage and Retrieval for Image and Video Databases III*, San Jose, CA (pp. 238-248).

Kumar, V. V., Reddy, B. E., Rao, A. N., & Raju, U. S. N. (2008). Texture segmentation methods based on combinatorial of morphological and statistical operations. [JMM]. *Journal of Multimedia*, *3*(1), 36–40. doi:10.4304/jmm.3.1.36-40

Li, J., Wang, J. Z., & Wiederhold, G. (2000). IRM: Integrated region matching for image retrieval. In *Proceedings of the 2000 ACM international conference on Multimedia* (pp.147–156).

Li, P., Guo, J., Song, B., & Xiao, X. (2011). 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*, *4*(1), 103–116. doi:10.1109/JSTARS.2010.2074186

Lucchese, L., & Mitra, S. K. (2001). Colour image segmentation: a state-of-the-art survey. *PINSA*, *67A*(2), 207–221.

Luo, B., Wang, X., & Tang, X. (2003). A world wide web based image search engine using text and image content features. In *Proceedings of the 2003 SPIE-IS&T Electronic Imaging, Internet Imaging IV* (pp. 123-130).

Ohashi, T., Aghbari, Z., & Makinouchi, A. (2003). Hill-climbing algorithm for efficient color-based image segmentation. In *Proceedings of the IASTED International Conference on Signal Processing, Pattern Recognition, and Applications (SPPRA 2003)*.

Pele, O., & Werman, M. (2009, September 27-October 4). Fast and robust earth mover's distances. In *Proceedings of the IEEE 12th International Conference on Computer Vision (ICCV 2009)*, Kyoto, Japan, (pp. 460-467).

Pentland, A., Picard, R. W., & Sclaroff, S. (1994). Photobook: Tools for content-based manipulation of image databases. In *Proceedings of the SPIE Conference on Storage and Retrieval for Image and Video Databases II, 2185,* San Jose, CA (pp. 238-248).

Prewer, D., & Kitchen, L. (2001). Soft image segmentation by weighted linked pyramid. *Pattern Recognition Letters*, *22*, 123–132. doi:10.1016/S0167-8655(00)00063-5

Rahimizadeh, H., Marhaban, M. H., Kamil, R. M., & Ismail, N. B. (2009). Color image segmentation based on Bayesian theorem and kernel density estimation. *European Journal of Scientific Research*, *26*(3), 430–436.

Rui, Y., Huang, T. S., & Chang, S.-F. (1997). Image retrieval: Past, present and future. *Journal of Visual Communication and Image Representation*, *10*, 1–23.

Sadek, S., & Al-Hamadi, A. Panning, A., Michaelis, B., & Sayed, U. (2009). A new approach to image segmentation via fuzzification of Rènyi entropy of generalized distributions. In *Proceedings of the World Academy of Science, Engineering and Technology* (Vol. 56, pp. 598-603).

Samadani, R., Han, C. S., & Katragadda, L. K. (1993). Content-based event selection from satellite images of the aurora. In *Proceedings of the SPIE Conference on Storage and Retrieval for Image and Video Databases, 1908,* San Jose, CA.

Schölkopf, B., Platt, J. C., Shawe-Taylor, J. C., Smola, A. J., & Williamson, R. C. (1999). Estimating the support of a high-dimensional distribution. *Microsoft Research Corporation Technical Report* MSR-TR-99-87.

Shao, H., Wu, Y., Cui, W., & Zhang, J. (2008, November 18–21). Image retrieval based on MPEG-7 dominant color descriptor. In *Proceedings of the 9th International Conference for Young Computer Scientists (ICYCS 2008)*, Zhang Jia Jie, Hunan, China (pp. 753-757).

Shih, F. Y., & Cheng, S. (2005). Automatic seeded region growing for color image segmentation. *Image and Vision Computing*, *23*, 877–886. doi:10.1016/j.imavis.2005.05.015

Stone, H. S., & Li, C.-S. (1996). Image matching by means of intensity and texture matching in the Fourier domain. In *Proceedings of the SPIE Conference on Storage and Retrieval for Still Image and Video Databases IV, 2670,* San Jose, CA.

Sumengen, B., & Manjunath, B. S. (2005). Multiscale edge detection and image segmentation. In *Proceedings of the 2005 European Signal Processing Conference (EUSIPCO 2005)*.

TRECVID. (2009). *Guidelines for the TRECVID 2009 evaluation, Section 5.3*. Retrieved from http://www-nlpir.nist.gov/projects/tv2009/tv2009.html

Vilaplana, V., & Marques, F. (2007). On building a hierarchical region-based representation for generic image analysis. In *Proceedings of the IEEE International Conference on Image Processing (ICIP 2007)*, San Antonio, TX (Vol. 4, pp. 325-328).

Wang, J. Z., Li, J., Chan, D., & Wiederhold, G. (1999). Semantics-sensitive retrieval for digital picture libraries. *D-Lib Magazine*, *5*(11). doi:10.1045/november99-wang

Wang, J. Z., Li, J., & Wiederhold, G. (2001). SIMPLIcity: Semantics-sensitive integrated matching for picture libraries. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *23*(9), 947–963. doi:10.1109/34.955109

Wong, K.-M., & Po, L.-M. (2004). MPEG-7 dominant color descriptor based relevance feedback using merged palette histogram. In *Proceedings of the Acoustics, Speech, and Signal Processing (ICASSP 2004)* (Vol. 3, pp. 433-436).

Xie, X., & Mirmehdi, M. (2007). Colour image segmentation using texems. *Annals of the British Machine Vision Association*, *1*(8), 1–10.

Xu, Y., Duygulu, P., Saber, E., Tekalp, A. M., & Yarman-Vural, F. T. (2000). Object based image retrieval based on multi-level segmentation. In *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP 2000)*, Istanbul, Turkey (Vol. 4, pp. 2019-2022).

Yu, Q., & Clausi, D. A. (2008). IRGS: Image segmentation using edge penalties and region growing. [PAMI]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *30*(12), 2126–2139. doi:10.1109/TPAMI.2008.15 PMID:18988947

Zhang, C., Zhou, L., Wan, W., Birch, J., & Chen, W.-B. (2010). An image clustering and feedback-based retrieval framework. [IJMDEM]. *International Journal of Multimedia Data Engineering and Management*, *1*(1), 55–74. doi:10.4018/jmdem.2010111204

Wei-Bang Chen is an Assistant Professor in the Department of Mathematics and Computer Science at Virginia State University. He also currently serves as a Research Scientist at Commonwealth Center for Advanced Manufacturing (CCAM). Dr. Chen earned his Doctoral degree as well as his Master's degree in computer science from the University of Alabama at Birmingham. He also holds a Master of Genetics from the Institute of Genetics, National Yang-Ming University, Taiwan. His research interests include multimedia data mining, image processing, computer vision, bioinformatics, network security, and manufacturing system automation.

Chengcui Zhang is an Associate Professor of computer and information sciences at the University of Alabama at Birmingham. Her research has been funded by Natural Science Foundation, eBay Inc., IBM, and NIH. She has authored and co-authored more than 120 research papers focusing in the areas of multimedia databases, multimedia data mining, bioinformatics, and geographic information systems. She is currently serving as a program co-chair of the 2013 IEEE International Conference on Information Reuse and Integration (IRI'13). She was a program co-chair for 2011 IEEE International Symposium on Multimedia and the 2012 International Conference on Information Reuse and Integration. She has served on more than 90 program committees of major conferences and workshops in multimedia and database systems. Zhang also serves on the editorial boards for several international journals, including International Journal of Multimedia Data Engineering and Management *and* International Journal of Computer and Informatics.