AN OBJECT-BASED APPROACH TO SIMILAR IMAGE RETRIEVAL

Andrea Kutics†‡, *Masaomi Nakajima*‡, *Toshiharu Ieki*‡

†Japan Systems Inc., Solution Division
2-31-24, Ikejiri, Setagaya-ku, Tokyo 154, JAPAN
‡NTT Data Corporation, Multimedia Technology Center
66-2, Horikawa-cho, Saiwai-ku, Kawasaki, Kanagawa 210, JAPAN
bs-andi@bs.mm.rd.nttdata.co.jp

ABSTRACT

This paper proposes a new object-based approach to similar image retrieval. First the salient regions of an image are detected by using a novel segmentation method based on a multi-scale inhomogeneous diffusion model applied to color and texture features. Each detected region is then represented by a feature vector composed from the characteristic color, texture and shape features of a region whose features are invariant under rotation. Scale invariance is also addressed for the color and shape properties. An R* tree-based indexing scheme is applied over the feature space to ensure efficient searching. By applying a suitable user interface, the method can handle sub-image queries and object-based queries with regard to a certain object or objects in the input image specified by the user. Experiments conducted on a large number of images taken from photo-CD data and collected from the Internet, show that the method performs well for a large variety of natural images.

1. INTRODUCTION

The rapid development of computer hardware, mass storage media and network devices, that has occurred in recent years, has resulted in explosive growth and wide accessibility of information in the field of electronic images and multimedia. In consequence, similar image retrieval has become essential in developing software tools capable of searching and extracting necessary information from the huge and ever-growing collection of image files appearing either on Internet pages or on other collection media. However, the task of retrieving similar images from large image collections is rather difficult to accomplish. It raises difficult questions, namely, how to effectively encode and represent the content of images, and how to provide appropriate similarity measures and indexing schemes. Additionally, there is the difficulty of providing a flexible interface enabling the user to present his/her specific queries, and fulfilling severe time requirements on producing similarity results within seconds. A number of different approaches to solving these problems have been proposed in recent years. First, traditional text indexing methods were utilized. However, due to shortcomings originating from the non-explicitness of image content and also from linguistic limitations, these approaches were soon replaced by methods utilizing visual properties. Most of these methods operate by extracting automatically attainable, low-level visual features which are closely related to and are assumed to be used in human visual perception. Then a suitable similarity measure is defined over the feature space to calculate differences between the query image and each database image. Global properties such as normalized color histograms that are invariant under geometric transformations and also texture features obtained by applying multi-channel filtering [1], SAR models [2], etc. are widely used as visual properties. In other approaches, geometric features, differential and moment invariant properties [3,4] are utilized. The advantage of using these latter properties, that they can capture shape-related features. More complex systems, such as MIT's Photobook [5] and Qbic [6] from IBM integrate multiple properties. However, the main drawback of using low level visual features obtained for the entire image is that it is difficult to provide an accurate description of the image content. It is assumed that human judgement on image similarity issues is largely influenced by shape and structural properties of recognized objects in the images. This suggests that more reliable methods can be achieved by integrating higher-level features that express structural and object-based information into the retrieval process. Work has just started on development of methods that utilize object-based information and carry out object-based queries, so we can mention only a few such methods. For example, the method in [7] uses ellipses to represent regions or blobs, while another method [8] integrates Bayesian color segmentation and spatial edges to generate regions. In developing systems using object-level features, the key issue is to provide an accurate segmentation algorithm to detect regions. This is extremely difficult to accomplish, as there is no restriction or preliminary knowledge available for the image domain on which the algorithm has to operate. The major difficulty in using object level features is that no general method has yet been provided for the problems of automatic image segmentation and object recognition.

This work proposes a new object-based approach to similar image retrieval. In this approach, first the salient regions of an image are detected by using a novel segmentation algorithm based on a multi-scale inhomogeneous diffusion model applied to color and texture features. Each detected region is represented by a feature vector composed from the characteristic color, texture and shape features of a region whose properties are invariant under rotation and scaling. An R* tree-based indexing scheme [9] is applied over the feature space to ensure efficient searching. We provide a user interface that enables the user either to carry out query-by-example searches or to search using his/her own example sketch. With this interface, the user can present a query not only by selecting an entire image but also by specifying a desired region or combination of regions in the selected or sketched image.

2. DETECTING SALIENT IMAGE REGIONS VIA AN INHOMOGENEOUS DIFFUSION MODEL

It is generally difficult to detect salient regions with quasi-uniform features within an image. This problem is inherently multi-scale due to the multi-resolutional characteristic of the human visual system. This work proposes an inhomogeneous diffusion model applied to multi-valued texture and single-valued color features to obtain a multi-scale framework. This model can be expressed as follows:

$$\frac{\partial F(x,y,t)}{\partial t} = div[d(x,y,t)grad\{F(x,y,t)\}] \quad (1)$$
$$F(x,y,0) = F_0(x,y)$$

where F indicates the feature vector and function $d(x,y,t)$ describes the diffusivity. We use the term inhomogeneous instead of anisotropic as it is not necessary to use different diffusivity functions for the two space variables, but rather a scalar valued diffusivity function analogously to the physical diffusion process taking place in an inhomogeneous multiphase medium. In this model, the color feature is simply represented by luminance values, while texture features are obtained by using a bank of Gabor filters [1] with a range of scale and orientation parameters. We have used three scales(s) and four orientations(o) in our experiments. The advantage of using multi-channel filtering is that the textural characteristic of the image can be analyzed on both local and global scales simultaneously. The choice of the family of Gabor filters is inspired by its close relation to low-level human visual perception, so that Gabor filters can model simple cells in the visual cortex. As Gabor filters are non-orthogonal, there is a certain overlap between the members of the generated filtering family, which can be minimized by applying the method proposed in [10].

We represent the texture features at each pixel position by feature vectors whose components are determined by calculating the energies of each filter in a small window (X,Y).

$$eng_{so}(x,y) = \frac{\sum_{X,Y} |GI_{so}(x,y)|}{(X*Y)} \quad (2)$$

where $GI_{so} = I(x,y) \otimes G_{so}(x,y)$ is a Gabor filtered image.

For both the texture and the color diffusion model, the diffusivity (d) is defined as a function of the corresponding feature gradient. For the texture features, we define a texture gradient at each image location (C_{ij}) as the maximum of the relative local differences of the texture feature components. This is given by:

$$C_{ij} = \max_{k \in \{T\}} \sum_{l,m} \left| \frac{T_{i,j}^{(k)}}{T_{\max}^{(k)}} - \frac{T_{i+l,j+m}^{(k)}}{T_{\max}^{(k)}} \right|$$
$$\{l,m \mid -1 \le l \le 1, -1 \le m \le 1, (l,m) \in \text{INT}, |l| + |m| > 0\} \quad (3)$$

where T_{ij} indicates a texture feature component(k) at a given image location. In this way, we detect texture boundaries by calculating relative changes in the local feature components, and these boundaries are not affected by the global texture energy. The diffusivity function [11] is defined by the following equation:

$$d(C) = 1/\left(1 + (C/K)^2\right) \quad (4)$$

where C indicates the feature gradient and K is a constant. The computation on the diffusion cycles of this multi-valued diffusion process is carried out in two main steps. First, the common diffusivity data are calculated by applying Eq. (3)-(4), and then the inhomogeneous diffusion steps are carried out separately for each component of the feature vectors by applying the following discretization model:

$$\frac{T_{i,j}^{n+1} - T_{i,j}^{n}}{\Delta t} = \frac{(d_{i+1,j} + d_{i,j})(T_{i+1,j}^{n} - T_{i,j}^{n}) - (d_{i,j} + d_{i-1,j})(T_{i,j}^{n} - T_{i-1,j}^{n})}{2(\Delta x)^2} + \frac{(d_{i,j+1} + d_{i,j})(T_{i,j+1}^{n} - T_{i,j}^{n}) - (d_{i,j} + d_{i,j-1})(T_{i,j}^{n} - T_{i,j-1}^{n})}{2(\Delta y)^2} \quad (5)$$

where T_{ij} indicates the texture features at a given image location, and d_{ij} is the diffusitvity at the same position. In the case of color features, a scalar-valued inhomogeneous diffusion process is carried out for the luminance values by applying a diffusivity which is a function of the luminance gradient. A multi-scale representation of the image can be obtained by carrying out diffusion cycles of increasing numbers over the texture and color features. First, we apply a boundary detection based on the diffusion results obtained at the different scale levels to detect multi-scale boundaries. Regions with quasi-uniform texture or color features are then obtained by applying an adaptive k-means algorithm tuned by the boundary information. Finally, a 'region growing and merging' step is applied to determine the salient object regions of the image (the size of each is no smaller than 1/64 of the image area). We can obtain characteristic texture and color features for the object regions from the center features of the clusters produced by the *k*-means algorithm. On the basis of the characteristic texture feature components, it can also be determined whether a region should be characterized as a texture region or as a color region. To ensure invariance under orientation changes, we calculate rotation-invariant texture features [12] from the characteristic texture features of the regions. These features are presented by a set of Fourier coefficients calculated for the texture feature components. Finally, we represent each salient image region by its characteristic color, texture, and shape features. The color features are expressed by the normalized color histograms (4 bins/axis) of the region, while the texture features are represented by a set of Fourier coefficients. The shape features of a region are expressed as a set of invariant properties, such as the normalized area of the region, the shape factor (circularity), the ratio of the major and minor axes, the normalized shape projections over the main axes and the normalized orientation histogram of boundary points (24 bins). The interconnections between the regions are also stored in a

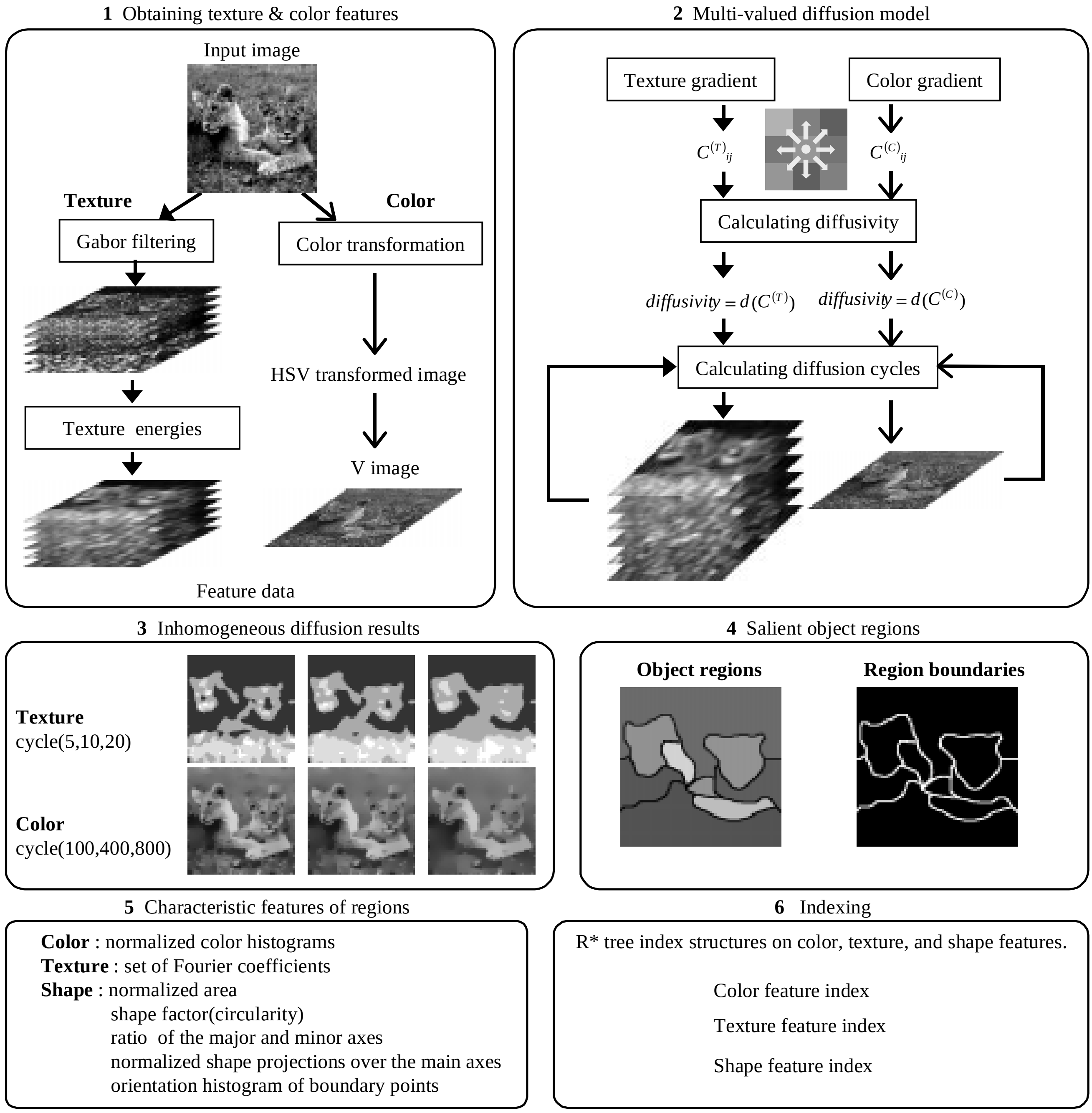
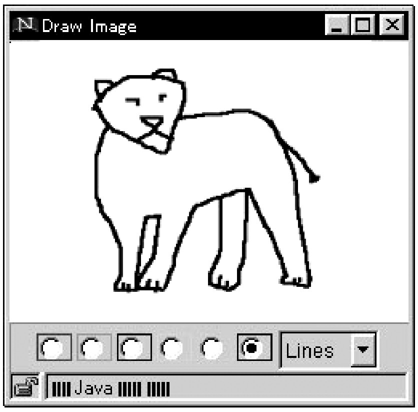


Fig 1. Outline of computation

region adjacency graph. The computation of the feature extraction and of the inhomogeneous diffusion model, as well as the object regions obtained for an image of a lion pair, is illustrated in Fig. 1. This figure also describes the characteristic features of the regions which are utilized to build up multidimensional index structures.

3. INDEX STRUCTURES AND PRESENTING QUERIES

In this work, an R* tree based indexing method is applied to support effective searching schemes. A feature vector is created from the color, texture and shape features for each detected region, and a multidimensional indexing algorithm is utilized to build up R* tree index structures using the feature vectors. In the R* tree, the distance between two feature vectors is calculated as a weighted sum of the Euclidean distances obtained for each component group of two vectors. We also build up R* trees independently for the color, texture and shape features to support independent color-, texture- or shape-based search schemes for nearest neighbor searches. In the indexes generated for searching on the basis of independent color or



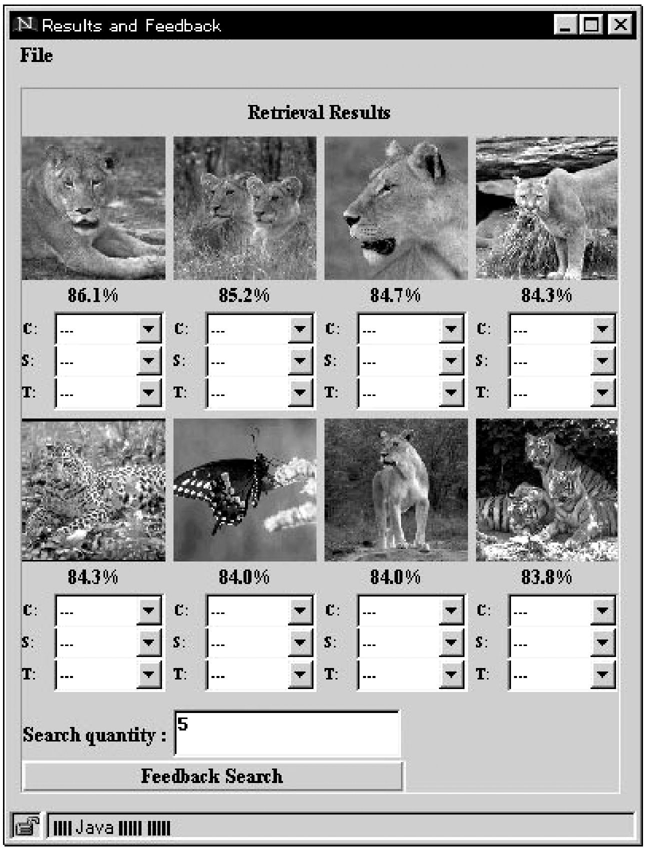


Fig. 2. Similarity results obtained for a simple object query (a: query frame showing the query image, the selected 'lion head' object and a segment of the database; b: the obtained similarity results; c: the drawing tool)

texture features, the sizes of the regions are also included. In the case of a query specifying a complex region – a region consisting of more than one simple region components – searches are also carried out for all of the region components. The final retrieval result in this case is determined by trying to reconstruct regions similar to the query region from the retrieved ones by using the region adjacency graph of the image. We provide a user interface that supports different types of user queries. The main interface is a query-by-example one, where the user can select a query image from the collection provided in a selection frame, or upload a query image of his/her own. The user is enabled to select a certain part or object in the query image by clicking the mouse over a point of the desired region when he/she wants to execute a search for only a certain object or objects. As a result of this action, the boundaries of the corresponding segmented area or areas are highlighted in red. There is also a ‘drawing-and-painting’ type interface. This additional tool is provided to enable the user to define an object region by sketching and painting it.

4. EXPERIMENTS

Experiments were conducted by applying the method to a photo-CD album containing 1200 photos of various types (landscapes, flowers, animals, etc.), and also on a set of more than 1000 images (photos, paintings, drawings, commercial images, etc.) collected from the Internet. We obtained an average retrieval rate of 82% on these two data collections for simple region queries and an average 74% rate for complex region queries. As anticipated, the obtained retrieval ratio was mostly influenced by the accuracy of the segmentation and object detection results. When the query image contains a larger number of small objects, or objects with low contrast and/or inhomogeneous texture and color features, the retrieval ratio decreases due to failures of the segmentation process. The method is implemented on a distributed computing platform consisting of four Sun workstations and several PCs. The main modules of the method, such as diffusion, segmentation, feature generation, indexing and searching modules are written as native C functions. These native functions are called from a multithreaded Java servlet-applet interface to produce search results within a few seconds. The main steps of the retrieval process and an example result are illustrated in Fig. 2. In this example, the query is presented as a ‘lion head’ object,specified by the user as a single particular object in a query image selected from the photo-CD collection. The similarity between the query image and the retrieved images is expressed as a similarity ratio with regard to the normalized distances in the feature space.

5. CONCLUSIONS

This paper has proposed a new object-based approach to similar image retrieval by representing an image by its salient objects as well as associated color, texture and shape features. We can summarize our achievements as follows. We have provided a novel and automatic segmentation method to quasi-accurately detect salient color and texture regions in the image. In this method, an inhomogeneous diffusion model is applied to both color and texture features. Characteristic color, texture and shape properties are provided for the detected regions that are invariant under rotation. Scale invariance was also addressed for the color and shape properties. By applying a suitable user interface, the method can handle sub-image queries, partial views and object-based queries with regard to a certain object or objects in the input image specified by the user. By enabling the user to specify the object he/she is looking for and accomplishing object-based searches, a higher level performance and flexibility is ensured. Experiments on a large number of images collected from the Internet and taken from photo-CD data showed that the method performs well for a large variety of natural images. In order to achieve better performance on images containing small object regions or regions with low contrast and/or inhomogeneous texture features, we need to improve the segmentation process. For this purpose, a better approximation of the multiphase diffusion process is considered necessary in the future by using a description of a parabolic diffusion equation with boundary conditions that depend on the feature gradient, and using a diffusivity calculated directly from the texture features.

REFERENCES

[1] A. K. Jain, N. K. Ratha and S. Lakshmahan, “Object detection using Gabor filters”, *Pattern Recognition*, 30(2), 295-309, (1997).

[2] J. Mao and A. K. Jain, "Texture classification and segmentation using multi-resolution simultaneous autoregressive models," *Pattern Recognition*, 25(2), 173-188, (1992).

[3] C. Schmid and Roger Mohr, "Local grayvalue invariants for image retrieval", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 19(5), 530-535, (1997)

[4] S. Siggelkow and H. Burkhardt, “Image retrieval based on color and nonlinear texture invariants”, *Proceedings of Noblesse Workshop of Nonlinear Model Based Image Analysis*, 217-224, (1998).

[5] A. Pentland, R. Picard and S. Sclaroff, "Photobook: Tools for content-based manipulation of image databases," *International Journal of Computer Vision*, 18(3), 233-254, (1996).

[6] M. Flickner,. et al, “Quaery by image and video content: The QBIC system” *IEEE Computer*, (23-30), (1995).

[7] C. Carson, S. Belongie, H. Greenspan and J. Malik, “Region-based image querying,” *Proceedings of CVPR '97 Workshop on Content-Based Access of Image and Video Libraries*, (1997).

[8] E. Saber, A. Murat Tekalp, “Integration of color, edge, shape, and texture features for automatic region-based image annotation and retrieval”, *Journal of Electronic Imaging* , 7(3), 684-700, (1998).

[9] N. Beckmann, H. P. Kriegel, R. Schneider and B. Seeger, “The R* tree: An efficient and robust access method for points and rectangles”, *Proceedings of ACM SIGMOD*, (1990).

[10] B. S. Manjunath and W. Y. Ma, “Texture features for browsing and retrieval of image data,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 8(11), 679-698, (1996).

[11] P. Perona, and J. Malik, “Scale-space and edge detection using anisotropic diffusion,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 12(7), 629-639, (1990).

[12] T. Sing Lee, “Rotation invariant texture features and their use in automatic script identification”, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 18(10), 751-756, (1998)