Content-based Image Retrieval with Intensive Signature via Affine Invariant Transformation

Timothy K. Shih, Lawrence Y. Deng, Ching-Sheng Wang, and Sheng-En Yeh
Dept. of Computer Science and Information Engineering Tamkang University
Tamsui, Taipei Hsien, Taiwan 251, R.O.C.
E-mail:tshih@cs.tku.edu.tw

Abstract

Web Content-based image retrieval is being implement in many databases especially in digital libraries. In many specialized fields, professionals often wish the content-based image retrieval system could provide more sensitive at shape and color matching to exploit or classify some kinds of species. In our experience, ecosystem synopsis of natural is the obvious illustration. Hence, we offered a series of methods to aim at these problems for solving. Such as defined the signature of the object, with Object Detection /Separation and Normalization, the Shape Representations and Similarity Measurement with affine invariants, the color feature and similarity measurement, the Content-based Image Retrieval System Requirements and Design issues.

Keywords: content-based image retrieval, signature, digital libraries

1. Introduction

With the growing popularity of World Wide Web, digital libraries over Internet play an important role in academic, the business, and the industrial worlds. Recent years have seen a rapid increase of size of digital image collections. Huge amount of information is out here. Image retrieval has been a very active research area since the 1970's. Because of the emergence of large-scale image collections, the manual annotation approaches become more and more difficult. That is, instead of being manually annotated by text-based keywords, images would be indexed by their own visual content, such as shape, color, texture, etc. Since then, many techniques in this research direction have been developed and many Image Retrieval Systems, both research and commercial, have been built. QBIC [7], is the first commercial content-based image retrieval system. QBIC supports queries based on example images, user-constructed sketches and drawings and selected color and texture pattern, etc. Its shape feature consists of shape area, circularity, major axis orientation and a set of algebraic moment invariant. Virge [10] is a content-based image search engine developed at Virage Inc. Similar to QBIC, Virage supports visual queries based on color, composition (color layout), texture, and structure (object boundary information). It also supports arbitrary combinations of the above four atomic queries. The users can adjust the weights associated with the atomic features according to their own emphasis. Photobook [8] is a set of interactive tools for browsing and searching images developed at MIT Media Lab. Photobook consists of three sub-books, from which shape, texture, and face features are extracted respectively. Users can then query based on corresponding features

0-7695-0933-9/00 $10.00 © 2000 IEEE

in each of the three sub-books. VisualSEEK [9] is virtual feature search engine, which is developed at Columbia University. Main research features are spatial relationship query of image regions and visual feature extraction form compressed domain.

In this paper, we will focus on shape and color feature. Because in many specialized fields, the professionals often wish the content-based image retrieval system could provide more sensitive at shape and color matching to exploit or classify some kinds of species. In general, the shape representations can be divided into to categories, boundary-based and region-based. The most successful representation for these two categories are Fourier Descriptor and Moment Invariant. The main idea of Fourier Descriptor is using the Fourier transformed boundary as the shape feature [1]. The main idea of Moment Invariant is using region-based moments, which are invariant to transformations, as the shape feature. Motivated by fact that most useful invariant was found by extensive experience and trial-and-error, Kapur et al. developed algorithms to systematically generates and search for a given geometry's invariant [2]. Some recent work in shape representation included Chamfer matching [3,4], Turning Function [5], and Wavelet Descriptor [6]. In this paper, we using the turning angle and affine invariant transformation model to matching the shape similarity.

There are common issues underlying all color-based retrieval methods: the selection of a proper color space [15], the use of a proper color quantization scheme to reduce the color resolution. Wang and Yang [16] reduce the color resolution by hierarchical clustering, CNS merging and equalize quantization method. Swain and Ballard [17] using histogram intersection as color indexing. Wan and Kuo [18] use hierarchical color clustering method based on the pruned octree data structure. For indexing the database, Pei and Shiue extract the color content of images and record the feature as index in [19]. In [20], Niu and Ozsu proposed an index scheme called 2D-h trees for content-based retrieval of images.

An object signature definition with Object Detection/Separation and Normalization are addressed in section 2. The Shape Representations and Similarity Measurement is discussed in section 3.The color feature is very important in content-based image retrieval domain, which is illustrated in section 4. And finally, we discuss our conclusion and possible extension in section 5.

2. Object Detection/Separation and Normalization

In order to provide the content-based image retrieval function into full play. It is necessary to preprocess the image for more easily to match the image's similarity with the media archive via network. In most situations, these images must be preprocessed via Object Detection (edge detection, edge tracing) and Object Normalization steps.

In this system, we divide the image and use the Seed Filling algorithm to extract the object. The procedure is described: Step1: Quantization and Normalization the image. Step2: Dividing the image. We divide an image into a number of boxes on the chessboard. The size of each box is 4*4 pixels and the representative color of a box is calculated by the average color of all pixels in the HSI color space. Step3: Filling the Seed. Starting from the upper-left corner, a box is chosen as a seed with the next box four units away both in the vertical and horizontal directions. Step4: Extract the objects. Starting from a seed box, the program looks at the left, right, up, and down directions. And, the program tries to combine as many boxes in a region as possible, if the color similarity between the seed box and the neighbor box is within a threshold.

The result of front steps will contain many regions (see Figure 2.1(c)). Each region contains some boxes. We remove the scattered small regions, because in general it is useless to retrieval the image and will decrease the performance (see Figure 2.1(d)). The procedure is describe follow:

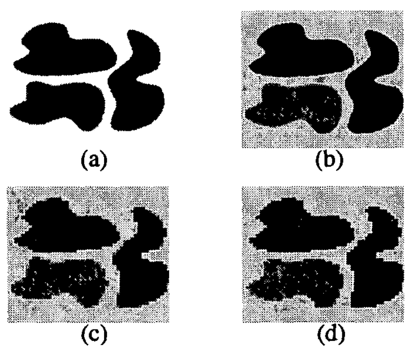

Fig. 2.1 Example of processed image (a) Original image.(b) Color clustering image. (c) Shape extraction image (d) Refined image

2.1 Object Detection

Edge detection and edge tracing are very important tasks in segmentation application of the images process system.

Technically, edge detection is the process of locating the edge pixels, and edge enhancement will increase the contrast between the edges and background so that the edges become more visible. In addition, edge tracing is the process of following the edges, usually collecting the edge pixels into a list. This is done in a consistent direction, either clockwise or counterclockwise around the objects. Chain coding [12] is one of the methods of edge tracing. The result is non-raster representations of the objects, which can be used to compute shape measurement or otherwise identity or classify the object.

2.2 Normalization

A signature is an important concept in image retrieval, and it can be define in a few different ways. We defined a signature by using orthogonal projection operator. Consider the image in Figure 2.2-1. The vertical profile (orthogonal projection on the X-axis) appears below as a weight graph. We can easily compute the middle line of the mass. The horizontal profile (orthogonal projection on the Y-axis) appears left side also same as below. This is easy to find the centroid (center of mass) of the object.

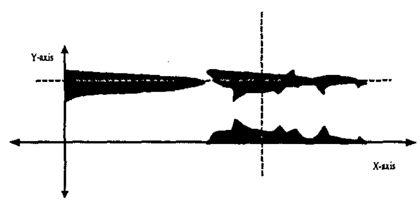

Figure 2.2-1 Use of orthogonal projection to find the centroid of the object.

After defined the centroid, the next step is to compute the longest distance form the centroid to the boundary. The longest distance form the centroid to the boundary can be extended to be “vital axis”. (as Figure 2.2-2 illustrated)

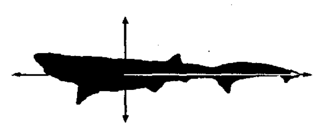

Figure 2.2-2 Representation of the object’s Vital axis.

The signature computed in this latter manner is dependent on rotation and scaling. Attempts have been made to normalize the signature to make it scale-independent.

3 Shape Representations and Similarity Measurement

Since shapes bear a semantic meaning, shape representation through a set of features, modeling prominent attributes of the shape, is the most popular technique. In this paper, we use the turning angle representation to describe the object. The turning angle of the object can express the edge's subtle difference including the curvature and distance. According the turning angle variation, we classify feature tokens based on their curvature properties. This method is more sophisticated.

We usually noticed distinguish between pre-attentive and attentive human similarity so that, we used the Affine-Invariant transformation method to be our similarity measurement model.

3.1 Shape Similarity Measurement with Affine-Invariant Transformation

In 1995, Wayne Niblack et al. proposed a pseudo-distance measure for 2D shape based on turning angle. In they report [11], through using turning angle to represent a shape, together with a dynamic programming algorithm to compute distance between them, gave the best overall results when compared with human perceptual rankings.

Considering the real world of the different camera view, it may occur that the user got the different view at the same object. Therefore, we discuss the deriving affine invariant[13,14] and novel shape similarity measurement model in this section. Let $(X_i, Y_i)\in S$ and $(X'_i, Y'_i) \in S'$ be the point pair on the original object and its affine object via the affine invariant transformation T= {[A], b}

$$\begin{bmatrix} X'i \\ Y'i \end{bmatrix} = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} \begin{bmatrix} Xi \\ Yi \end{bmatrix} + \begin{bmatrix} b_1 \\ b_2 \end{bmatrix} \quad ,(i=1,2,\ldots n),$$

(3.3.1)

where $[A] = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}$, translation is represented by $\begin{bmatrix} b_1 \\ b_2 \end{bmatrix}$,

Lemma 1: If (X_s, Y_s) is the starting point of S, and and (X'_s, Y'_s) be the starting point $\in S'$ pair on the original object and its affine object via the affine invariant transformation T= {[A], b}

$$\begin{bmatrix} X's \\ Y's \end{bmatrix} = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} \begin{bmatrix} Xs \\ Ys \end{bmatrix} + \begin{bmatrix} b_1 \\ b_2 \end{bmatrix} ,(i=1,2,\ldots n), \quad (3.3.2)$$

Proof : Owing to the (X_s, Y_s) is the starting point of S, so $(X_s, Y_s) \in S$, And and (X'_s, Y'_s) be the starting point $\in S'$ pair on the original object and belonging to S', by equation (3.3.1), s.t. its affine object via the affine invariant transformation T= {[A], b}

$$\begin{bmatrix} X's \\ Y's \end{bmatrix} = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} \begin{bmatrix} Xs \\ Ys \end{bmatrix} + \begin{bmatrix} b_1 \\ b_2 \end{bmatrix} ,(i=1,2,..n),$$

Lemma 2: If (X_c, Y_c) is the centroid of S, and and (X'_c, Y'_c) be the centroid point $\in S'$ pair on the original object and its affine object via the affine invariant transformation T= {[A], b}

$$\begin{bmatrix} X'c \\ Y'c \end{bmatrix} = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} \begin{bmatrix} Xc \\ Yc \end{bmatrix} + \begin{bmatrix} b_1 \\ b_2 \end{bmatrix},$$

i=1,2,…n), (3.3.3)

Proof : Owing to the point (X_c, Y_c) is the centroid of S is the average values of all the points (X_i, Y_i) belonging to S, so as the (X'_c, Y'_c). It is obvious to show that the (X'_c, Y'_c) also obey the the affine invariant transformation T= {[A], b}

$$\begin{bmatrix} X'c \\ Y'c \end{bmatrix} = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} \begin{bmatrix} Xc \\ Yc \end{bmatrix} + \begin{bmatrix} b_1 \\ b_2 \end{bmatrix}, (i=1,2,\ldots n),$$

Form equation (3.3.3) and (3.3.1), we can derive the relative invariant:

$$[(X'_i - X_c)(Y'_i - Y_c) - (X_i - X_c)(Y_i - Y_c)] =$$

$$det\{[A]\}\,[(X_i - X_c)(Y_i - Y_c) - (X_j - X_c)(Y_j - Y_c)] \quad (3.3.4)$$

By equation (3.3.4),we can find the affine shape have an area relative invariance property that shows the area S of an affine transformed object is equal to the product of its original object area S times the determinant of the transformation matrix:

$$S' = det\{[A]\}S \qquad (3.3.5)$$

Let θ be the original turning angle vector. By equation (3.3.5), we can derive a geometric interpretation:

$$\theta' = det\{[A]\}\ \theta . \qquad (3.3.6)$$

Let θ'_1 be the affine transformed form θ_1. It computes the global best match between θ'_1 and θ_2 in the sense that it pairs up each element of θ'_1 with an element of θ_2 to give the minimum sum of absolute differences. In the pairing, multiple elements of θ'_1 can match a single element of θ_2 (and vice versa), but the matching must proceed monotonically through both sets. Thus it computes an angle index pair sequence $(i_1,j_1), (i_2,j_2),\ldots(i_k,j_k)$, such that either $i_{t+1}= i_t$ or $i_{t+1}= i_t +1$ (similarly for i), by minimizing the distance between matched turning angle points turning :

$$D= \sum_{t=1,\ldots,k} |\det\{[A]\}(\theta_1(i_t) - \theta_2(j_t))| + r \quad , \quad (3.3.7)$$

Where r=0 if $i_{t+1} = i_t + 1$ and $j_{t+1} = j_t + 1$; Otherwise, it attaches a "warping" penalty.

Let the similarity measurement degree from 1(nearest matching) to 0(most dissimilar matching). The measurement of between the requested image and archive images is:

$D_a = 1 - \dfrac{MinD_i}{MaxD_k}$, where D_k is the maximum distance measure of the requested image and archives images, D_i is the minimum distance measure of the requested image and archives images.

If the archives images include the requested image, then the $MinD_t$ is equal to 0. The D_a is equal to 1. In addition, the $MaxD_k$ can set a big constant coefficient (it depends on the turning angel number). If the $MinD_t > MaxD_k$ means the requested image is dissimilar to archives images, , then we set the D_a to be 0.

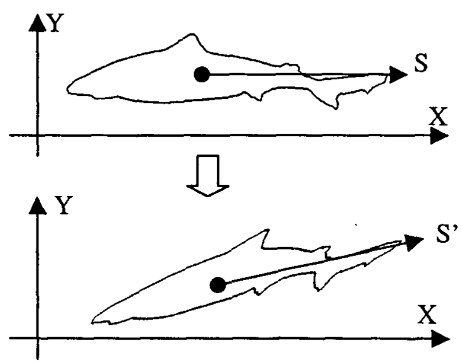

Figure 3.2 Shape relative affine invariants transformation with starting point.

4. The Selection of Color Space

The selection of a proper color space and the use of a proper color quantization scheme to reduce the color resolution e are common issues underlying all color-based retrieval methods [15,16]. A color space is a mathematical representation of a set of colors. There are several color spaces existing for a variety of reasons.

In our study, HSI is the employed color space because of its similarity and perceptibility. Similarity means that two perceptually similar colors will be in the same or neighbor quantized color bin and two non-similar colors will not be in the same quantized color bin, so the similarity of two colors can be determine according to the distance in HIS color space. Besides, the HIS color space is defined based on the human color perception, so user could choice the color he/she wanted easily by indicate hue, saturation, intensity value independently. In addition, user can express the query based on value of hue, saturation and intensity to select the sensation of image between warm and cold, saturation and instauration and bright and dark.

4.1 Color Clustering and Indexing Scheme of Image Database

In this section, we present our mechanism and procedure of color clustering and the normalization of image, involve MTM transformer formulas from RGB to HIS color space. In addition, the index scheme and filer mechanism according to clustering scheme and human sensation for speeding up the retrieval process will be described.

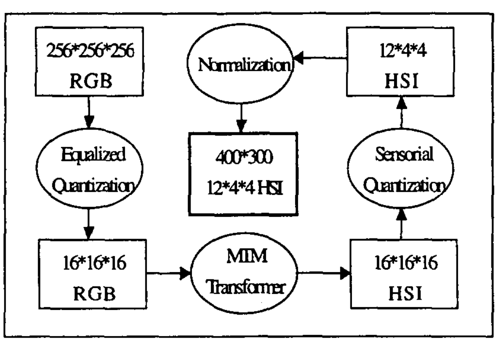

Fig.4.1 Procedure diagram of the color clustering and normalization

Color Clustering and Normalization

The quantization scheme and the procedure of color clustering are illustrated in Figure 4.1. Firstly, we equally quantized the RGB color space to change color levels from 256 to 16 levels in each axis. Secondly, we linearly convert the 16-level RGB color bins to the HSI coordinates by MTM transformer formulas:

$$H = \arctan 2(\sqrt{3}(G-B),(2R-G-B))$$

$$S = 1 - \frac{\min(R,G,B)}{I}$$

$$I = \frac{(\mathrm{R}+\mathrm{G}+\mathrm{B})}{3}$$

And, we cluster the hue to 12 levels, since hue is represented as circle and primary hues are located on the equal space at 60 degrees (Red, Yellow, Green, Cyan, Blue and Magenta) in the HSI color space. And, because the human visual system is more sensitive to hues as compared to saturation and intensity, the H axis should be quantized finer than S axis and I axis. In experimenting, we quantized the HSI color space into 12 bins for hue, 4 bins for saturation, and 4 bins for intensity (Figure 4.2). Finally, We normalize the resolution of all images to be 400*300.

The Indexing Scheme

After the quantization and normalization, system will index the images according to the dominant colors of those images. First, system will calculate the histogram and dominant colors of the image. The color histogram is an array that is computed by differentiating the colors within the images and counting the number of pixels of each color. From the color histogram, we could choose the dominant colors whose numbers of pixels exceeds the threshold.

After getting the dominant colors, system will save the unique image ID to each corresponding color bin. And, the logical indexing address which length equals to 1 byte only (0100 01 01=69) of each color bin in the database could be obtained immediately according to their hue (0100), saturation (01), and intensity (01).

The Filter Mechanism

For a small image database, sequentially searching the image during the retrieval process will be fast and provide acceptable response time. However, it is not feasible for large image database. Therefore, we propose a filtering mechanism to eliminate irrelevant images before the more complex and expensive similarity measure.

First, system will load the image ID arrays according to the dominant colors of query image. Next, system will conjunc and rank the image ID arrays according to the number of appearance. For example: If the dominant colors of an image are 69(01000101), 70(01000110), 71(01000111), 184(10111000), 186(10111010). Then the image ID arrays of those color bins are 1,2,3; 1,2,8; 2,7,9; 2,3,7,9 and 1,2,9 will be loaded (show in Figure 3). And, because image ID 2 appear 5 times, image ID 1 and 9 appear 3 times, image ID 3 and 7 appear twice and image ID 8 appear only once, the result of conjunction and ranking those arrays are: 2,1,9,3,7,8. After this step, system could filter out the irrelevant images effectively (ex: 4,5,6,10...).

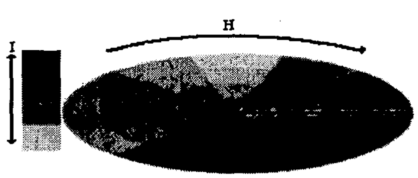

Figure 4.2. Clustering 12*4*4 HSI

In addition to the filter mechanism described above, system could filter the image based on the color sensation of human beings. According to the chromatology, the hue value influences the sense of warm or cold, the saturation value influences the sense of saturated or unsaturated and the intensity value influences the sense of bright or dark. When user indicates the color-sensation query about those three senses by him/herself or analysis the input image by the system, filter could load relevant image IDs from our special HSI indexing structure efficiently.

5. Conclusions and Future Work

In this paper, we proposed a series of methods to solve the image retrieval problems especially in shape and color similarity. Such as defined the signature of the object, with Object Detection/Separation and Normalization, the Shape Representations and Similarity Measurement with affine invariants, the color feature and similarity measurement, the Content-based Image Retrieval System Requirements and Design issues. We also provided the integrated content-based retrieval system service that contains the subsystem for the extraction of perception features of visual data, an index structure and a relevance feedback mechanism. However, visual information retrieval is a important subject of research in information technology. Our future work will integrate other methods such as spatial relationship or texture etc. for searching the object from the image directly, professionally and efficiently.

6. References

[1] E. Person and K. S. Fu, "Shape discrimination using Fourier Descriptors," IEEE Trans. Sys. Man, Cyb., 1977.

[2] D. Kapur, Y. N. Lakshman, and T. Saxena, "Computing invariants using elimination

methods," in Proc. IEEE ,Int. Conf. on Image Proc., 1995.

[3] H. G. Barrow, "Parametric correspondence and chamfer matching: Two new techniques for image matching," in Proc 5th Int. Joint Conf. Artificial Intelligence, 1977.

[4] G. Borgefors, "Hierarchical chamfer matching: A parametric edge matching algorithm," IEEE Trans. Patt. Recog. and Mach. Intell., 1988.

[5] E. M. Arkin, L. Chew, D. Huttenlocher, K. Kedem, and J. Mitchell, "An efficiently computable metric for comparing polygonal shapes," IEEE Trans. Patt. Recog. and Mach. Intell., vol. 13, March 1991.

[6] G. C.-H. Chuang and C.-C. J. Kuo, "Wavelet descriptor of planar curves: Theory and applications," IEEE Trans. Image Proc., vol. 5, pp. 56-70, January 1996.

[7] M. Flickner, H Sawhney, W. Niblack, J. Ashley, Q Huang, B. Dom, M. Gorkani, J Hafine, D. Lee, D. Petkovic, D. Steele, and P. Yanker, "Query by Image and Video Content: the QBIC system," IEEE Computer, 1995.

[8] A. Pentland, R. Picard, and S. Sclaroff, "Photobook: Content-based manipulation of image databases," International Journal of Computer Vision, 1996.

[9] J. R. Smith and S.-F. Chang, "Visualseek: A fully automated content-based Visual Query System," in Proc. ACM Multimedia 96, 1996.

[10] A. Gupta and R. Jain, " Visual information retrieval," communications of the ACM, vol. 40, no. 5, 1997

[11] *Niblack, W.; Yin, J.* Image Processing , "A pseudo-distance measure for 2D shapes based on turning angle" International Conference on Volume: 3 , Page(s): 352 -355 vol.3,1995. Proceedings.

[12] Freeman, H., " Computer processing of line drawing images", Computer Survey, 6:57-97, 1974.

[13] Zhengwei Yang and Fernand S. Cohen, "Cross-Weighted Moment and Affine Invariants for Image Registration and Matching", IEEE Transaction on Pattern Analysis and Machine Intelligence, Vol. 21, No. 8, 1999.

[14] J. Mundy and A. Zisserman, "Geometric Invariance in Machine Vision",ed. MIT Press, 1992.

[15] Xia Wan and C.-C Jay Kuo, "Color Distribution Analysis and Quantization for Image Retrieval," in SPIE storage an Retrieval for Image an Video Database IV, Vol. SPIE 2670, Feb. 1996, pp.9-16.

[16] Jia Wang, Wen-jann yang and Raj Acharya, "Color Clustering Techniques for Color-Content-Based Image Retrieval from Image Databases," IEEE International Conference on Multimedia Computing and Systems, ICMCS Jun 3-6. 1997, pp.442~449.

[17] M.J. Swain and .H. Ballard, "Color index," Internal Journal of Computer Vision, Vol.7, No.1, 1991, pp.11~32.

[18] Xia Wan and C.-C Jay Kuo, "A New Approach to Image Retrieval with Hierarchical Color Clustering," IEEE Transaction on Circuit and Systems for Video Technology, Vol.8, No.5, September 1998, pp.628~643.

[19] Soo-Chang Pei an Hui-Lirng Shiue, "Indexing and Retrieval of Color Image Database," pp.196~199.

[20] Youping Niu, M. Tamer Ozsu and Xiaobo Li, "2D-h Trees: An Index Scheme for Content-Base Retrieval of Image in Multimedia Systems," IEEE International Conference on Intelligent Processing Systems, October 28-31. 1997, pp.1710~1715.