INTELLIGENT RETRIEVAL
Open Journal

Review

An Intelligent Content-based Image Retrieval System Based on Color, Shape and Spatial Relations

Timothy K. Shih*; Jiung-Yao Huang; Ching-Sheng Wang; Jason C. Hung; Chuan-Ho Kao

Department of Computer Science and Information Engineering Tamkang University, Tamsui, Taiwan, R.O.C.

****Corresponding author***
Timothy K. Shih
Department of Computer Science and Information Engineering, Tamkang University, Tamsui, Taiwan, R.O.C.

Article information
Received: July 30th, 2024; **Revised:** August 30th, 2024; **Accepted:** September 22nd, 2024; **Published:** October 4th, 2024

Cite this article
Shih TK, Huang J-Y, Wang C-S, Hung JC; Kao C-H. An intelligent content-based image retrieval system based on color, shape and spatial relations. 2024; 2(2).
doi: https://doi.org/10.70705/ppp.ir.2024.v02.i02.pp82-87

ABSTRACT

Research into content-based multimedia information retrieval is both fascinating and challenging. The use of shape, texture, and color data is one example of a modern technique. In this study, we provide a hybrid method that has been used to recover images from around 5,000 photographs; this method takes into account the color, form, and spatial interactions among items in a picture. Based on our clustering approach and color sensibility, we use a redesigned color scheme and its indexing strategy to enhance retrieval efficiency. The process of extracting form and spatial connections from objects is akin to seed filling. When comparing geographical disparities for similarity, a qualitative technique is used. In addition to a user-friendly graphical user interface (GUI), the system incorporates sketch pictures and relevance feedback to enhance retrieval accuracy. Our results demonstrate that the system can effectively extract picture data using the suggested method.

Keywords
Content-based image retrieval; Image database; Color; Shape; Spatial relation.

INTRODUCTION

Text annotations are used by the majority of traditional picture databases. Therefore, searching for keywords is the foundation of image retrieval. Adding text to photographs makes them simpler to work with and manage. On the other hand, this approach has two big flaws. One thing to keep in mind is how much time it takes to create keywords for a lot of photographs. Not to mention that the keywords are subjective and not unique by definition. These drawbacks make content-based retrieval and automated indexing more appealing for building large-volume picture retrieval systems.

When it comes to picture recognition, both humans and computers rely on color, form, and spatial interactions. In order to improve content-based information retrieval, many approaches have been suggested by researchers in the last few years:

(1) QBIC (Query By Image Content) by IBM,
(2) Virage by Virage, Inc.,
(3) Photobook by MIT Media Lab.,
(4) VisualSEEK by Columbia University,
(5) RetrievalWare by Excalibur Technologies Corp.,
(6) NeTra by the University of California, Alexandria Digital Library,
(7) IRIS by German Software Development Laboratory of IBM and the AI group of the University of Bremen,
(8) CORE by the University of Singapore.

QBIC was the first commercial CBIR (Content Based Image Retrieval) system actually built (Flicker et al., 1995).

Its methods and system architecture have impacted subsequent image retrieval systems. Color, form, texture, sample photos, and drawings are some of the ways the QBIC system searches its extensive picture and video collection. The geometric transform to Munsell (MTM) coordinates, the average (R,G,B), the (Y,i,q), and the (L,a,b) color characteristics are used in QBIC.

According to Bach et al. (1996), the Virage Search Engine allows users to query both still images and video streams. It allows visual searches based on structure (object boundary information), texture, color pattern (composition), and color, much like QBIC. The four atomic queries shown above may be combined in any way you choose. The customer has the option to customize the atomic feature weights to suit their needs.

Interactive tools for viewing and finding photographs make up the Photobook system (Pentland et al., 1996). The core principle of these technologies is semantics preserving image compression, which condenses a picture to a handful of coefficients that are still noticeable to the human eye. The photobook is divided into three sections. A person's face, form, or texture may be extracted using the

Appearance Photobook, form Photobook, or Texture Photobook, respectively. A user may use one or more of the three sub-books' relevant characteristics to query a picture, or they can combine several processes with a textual description. Alshuth et al. (1996), Dowe (1993), Ma and Manjunath (1997), Wu et al. (1995), and other systems (i.e., 3, 6, 7, and 8) provide querying based on global color, color scheme, form, texture, and semantic content. Image region spatial relationship searching is the primary

visual search engine (Smith and Chang, 1996). It gives you the ability to search using criteria like color, texture, and spatial arrangement.

No content-based image retrieval system combines global color, color region, color sensation, form, and qualitative spatial connection characteristics for picture querying, even though all of them provide numerous attributes for image retrieval. In this article, we'll go over those characteristics for picture feature extraction and measurement. After then, the issue is resolved in its entirety.

The choice of an appropriate color space and the implementation of an appropriate color quantization technique to decrease color resolution are fundamental concerns of all color-based retrieval systems. Through the use of hierarchical clustering, CNS (Color Name System) merging, and an equalization, quantization approach, Wang et al. (1997) were able to decrease color resolution. When it came time to index colors, Swain and Ballard (1991) turned to histogram intersection. In this study, Wan and Kuo (1996) used a pruned octree data structure–based hierarchical color clustering approach. Using the HIS color space, we group color resolutions based on how they feel to the human eye in this research.

Given that humans have a tendency to disregard such differences when it comes to identification and retrieval, it is crucial that form characteristics remain same when subjected to translation, rotation, and scaling. You may classify shape representations as either boundary-based or region-based. The Fourier Descriptor and the Moment Invariant are the best representations of these two classes. The Fourier Descriptor relies on the shape characteristic that is the Fourier converted border (Person and Fu, 1977). Making advantage of region-based moments—which are shape features that are invariant to transformation—is fundamental to the Moment Invariant. Kapur et al. (1995) created algorithms to systematically generate and search for an invariant of a given geometry, inspired by the fact that the best invariants are discovered by hand and via trial and error. Chou and Kuo (1996), Arkin et al. (1991) and the Chamfer Matching (Barrow, 1977; Borgefors, 1988), as well as the Wavelet Descriptor (Arkin et al., 1991). Image retrieval issues, such as shape representation and matching, object signature definition, statistical identification, and species categorization, are addressed in this work via the proposal of a number of approaches.

Multimedia object temporal modeling has been suggested by several scholars. In addition to 1D and 2D objects, the mechanism can handle 3D objects that use the R-tree spatial data structure (Vazirgiannis et al., 1996) and use a time line as the third axis. A number of kinds of temporal interaction were discussed in Wahl et al. (1995), along with the temporal relationships between them. Some interaction types were also included into a temporal model in that work. According to Nabil et al. (1996), the 2D Projection Interval Relations method (PIR) makes use of topological and directional temporal relations. We spoke about PIR-based image retrieval techniques. Many multimedia applications deal with the underlying representation of objects via the use of spatio-temporal relations, which provide a reasonable semantic tool. In this work, we use full analysis for spatial computing to expand temporal interval connections.

Furthermore, an image database's index strategy is crucial for better retrieval performance. An indexing technique based on binary trees was created by Visual-SEEK (Smith and Chang, 1996). Histogram intersection was used as a color indexing method by Swain and Ballard (1991). The characteristic was documented as an index by Pei and Shiue (1998) after they extracted the color components of photographs. For picture retrieval based on content, Niu et al. (1997) suggested an index technique known as 2D-h trees. Using our clustering approach in conjunction with MTM, we provide a fast database indexing strategy that can efficiently remove photos that are not related.

Here is the structure of the remaining portion of our article. Section II covers a few of color spaces, such as RGB, CMY, YUV,

In order to choose the optimal color space for color clustering, we examine and compare C.I.E. L*u*v* and HSI. We talk about an image database's color clustering and indexing technique in Section III. In Section IV, we detail the feature extraction method. Section V provides the similarity functions for the geographic relation characteristics, color, and form. In Section VI, we detail how our system retrieves images. Section VII concludes the paper briefly.

Selection of the Color Space

Selection of a proper color space and use of a proper color quantization scheme to reduce the color resolution are common issues for all color-based retrieval methods (Swain and Ballard, 1991; Wan and Kuo, 1996; Wang et al., 1997). A color space is a mathematical representation of a set of colors. Several color spaces exist for a variety of reasons. Some color spaces are widely used color spaces in computer graphics.

1. RGB Color Space

Red, green, and blue are the three primary additive colors (individual components are added together to form a desired color) and are represented in three dimensions. The RGB color space is the most prevalent choice for digital images because color-CRTs (computer display) use red, green, and blue phosphors to create the desired color. Also, it is easy for pro- grammers to understand and program since this color space has been widely used for a number of years. However, a major drawback of the RGB space is that it is senseless. The user finds it difficult to understand or get a sense of what color R = 100, G = 50, and B = 80 is and the difference between R=100, G = 50 and B = 50, and R = 100, G=150 and B = 150.

2. CMY Color Space

The CMY color space is used for color printing. Cyan, Magenta, Yellow are the complements of Red, Green and Blue. They are called subtractive primaries because they are obtained by subtracting light from white. RGB to CMY can be obtained by means of a simple, but inaccurate transformation: C = 1

 R, M = 1 G and Y = 1 B. Hence, the CMY color space has the same drawback as RGB.

3. C.I.E. L*u*v* Color Space

A color is identified by two coordinates, x and y, in the

C.I.E. L*u*v* Color Space. Lightness L* is based on a perceptual measure of brightness, and u* and v* are chromatic coordinates. Also, color differences in an arbitrary direction are approximately equal in this color space. Thus, the Eu- clidean distance can be used to determine the relative distance between two colors. However, co-ordinate transformation to the RGB space is not linear.

4. YUV, YIQ and YCbCr Color Space

The YUV space is widely used in image compression and processing applications. Y represents the luminance of a color, while U and V represent the chromaticity of a color. The luminance (Y) component is separated from the chromatic components in this space. The YIQ color space is derived from the YUV color space. I stands for In-phase and Q for Quadrature, which is the modulation method used to transmit color information. YCbCr is a scale and offset version of the YUV color space. These color spaces are difficult for users to deal with because they do not directly refer to intuitive notions of hue, saturation and brightness.

5. HSI, HSV and HLS Color Space

The HSI color space was designed to be used more intuitively in manipulating color and to approximate the way humans perceive and interpret color. Three characteristics of color, hue, saturation and intensity, are defined so as to dis- tinguish color components. Hue describes the actual wave- length of a color by representing the color's name, for example, green, red or blue. Saturation is a measure of the purity of a color. For instance, the color red is a 100% saturated color, but pink is a low saturation color due to the a mount of white in it. Intensity indicates the lightness of a color. It ranges from black to white. HLS (hue, lightness, and saturation) is similar to HSI; the term lightness is used rather than intensity. The different between HSI and HSV (hue, saturation, value) lies in the computation of the brightness component (I or V), which determines the distribution and dynamic range of both brightness (I or V) and saturation (S).

In our study, HSI is the color space employed because in terms of its resemblance and discernibility. What we mean by "similarity" is that two colors that are visually close are in the same or nearby quantized color bins, but two colors that are visually different are not in the same bin. So, the distance in the HIS color space is a good indicator of how similar two colors are. Human visual perception also forms the basis for the HIS color space. By individually designating the hue, saturation, and intensity levels, users may simply pick the color they want. Users may also convey their visual sensations—such as warmth or coolness, grayscale or vividness, or brightness or darkness—by querying depending on saturation and intensity levels.

II. Color Clustering and Indexing Scheme for an Image Database

In this section, we present our mechanism and procedure for color clustering and normalization of images, including the MTM transformer formulas obtained from the RGB to HIS color space. In addition, the index scheme and filter mechanism, based on the clustering scheme and human sensation, for speeding up the retrieval process are described.

1. Color Clustering and Normalization

The quantization scheme and the procedure for color clustering are shown in Fig. 1. Firstly, we equally quantize the RGB color space to change the number of color levels from 256 to 16 levels in each axis. Secondly, we linearly con- vert the 16-level RGB color bins into HSI coordinates using the MTM transformer formulas:

on the circle every 60 degrees (Red, Yellow, Green, Cyan, Blue and Magenta) in the HSI color space. In addition, because the human visual system is more sensitive to hues than to saturation or intensity, the H axis is quantized more finely than the S axis and the I axis. In our experiments, we quantized the HSI color space into 12 bins for hue, 4 bins for saturation, and 4 bins for intensity (Fig. 2). Finally, we normalized the resolution of all the images to 400 □ 300 pixels.

2. The Indexing Scheme

After quantization and normalization, our system in- dexes the images according to their dominant colors. First, the system calculates the histogram and dominant colors of the image. The color histogram is an array that is computed by differentiating the colors within the image and counting the number of pixels of each color. From the color histogram, we can choose the dominant colors whose numbers of pixels exceed the threshold.

After getting the dominant colors, the system saves the unique image ID to each corresponding color bin. For example, if the unique ID of an image is 1, and the dominant colors of this image are 37 (00100101), 38 (00100110), and 154 (10011010), then the image ID is saved to the ID arrays in the logical indexing addresses 00100101, 00100110, and 10011010 (Fig. 3). According to this indexing scheme, the system can load the candidate images that have same domi- nated colors and eliminate irrelevant images immediately before the more complex and expensive similarity measure- ment is carried out.

3. The Filter Mechanism

A tiny picture database may achieve quick and acceptable response time by searching each image sequentially during retrieval. Nevertheless, a vast picture database will not be able to handle this. Consequently, we provide a filtering approach to remove superfluous photos prior to running the more involved and costly similarity test. Our method begins by loading the picture ID arrays based on the query image's prominent colors. The next step is for the system

III. Extraction of Features

Here we showcase our algorithms that can extract picture characteristics. Common picture formats include MP, GIF, JPEG, and many more. Bitmaps are the foundation of our feature extraction approach since the spatial information they provide is inherent in their pixel streams. It is only during the feature extraction process that images saved in other formats are converted to bitmap format. Following feature extraction, the system records both the extracted feature data and the original picture format in a database. To summarize, our system is capable of processing images in various formats as long as they are converted to bitmap during the feature extraction step.

1. Color Feature Extraction

There are primarily two steps to the process of color feature extraction. In order to get the HSI value, the system reads the RGB values of the pixels in the picture file in a sequential fashion and then uses the MTM transformer to convert them. Second, the system builds the image's color histogram according to the pixel counts of each color.

Also, the system records the color sensation of the image, including warmth-coldness, gray-vividness and brightness- darkness, according to the average hue value, saturation value and intensity value of the image for filtering.

2. Shape Feature Extraction

In form extraction, you'll need to do three things: We begin by applying the improved Seed Filling algorithm to the picture in order to extract its forms. Prior to measuring shape similarity, we use edge detection to normalize the forms and transform them from regions to contours. To begin, the system finds the color and spatial connections between the items and records their position, height, breadth, area, and color for use in similarity measuring. Only for the purpose of measuring form similarity is the normalized contour, which is created in stages two and three, used.

A. Shape Extraction

The procedure in the Seed Filling algorithm for extract- ing region features is described below:

Step 1: Quantization and normalization of the image.

Step 2: Dividing the image. We divide the image into a number of boxes on a chessboard. The size of each box is 4 □ 4 pixels, and the representative color of a box is calculated based on the average color of all the pixels in the HSI color space.

Step 3: Filling the seed. Starting from the upper-left corner, a box is chosen as a seed with the next box four units away, in both the vertical and horizontal directions.

Step 4: Extract the objects. Starting from the seed box, the program looks in the left, right, up, and down directions. Then, the program tries to combine as many boxes in a region as possible if the color similarity between the seed box and the neigh- boring box is within a threshold.

The result of Steps 1 □ 4 will contain many regions (Fig. 4(c)). Each region will contain some boxes. We remove the scattered small regions because, in general, these regions are useless in image retrieval and will worsen performance (Fig. 4(d)).

B. Edge Detection

Segmentation schemes in image processing systems rely heavily on edge detection and tracing. The Canny operator and the Shen-Castan (ISEF) approach are two popular signal edge detectors (Canny, 1986; Shen and Castan, 1992). The Canny algorithm applies non-maximum suppression and hysteresis thresholding after convolving the image with the derivative of a Gaussian function; the Shen-Castan algorithm computes the binary Laplacian image, suppresses false zero crossings, performs adaptive gradient thresholding, and applies the hysteresis threshold; and the Canny algorithm uses the Infinite Symmetric Exponential Filter (ISEF) to convolve the image. Our research indicates that Canny's edge detector performs well in a wide variety of specific contexts. To make sure the boundary line is narrow and no wider than one pixel, non-maximum suppression is used.

C. Representing and Normalizing Shapes

Locating edge pixels is the technical definition of edge detection, whereas tracking the edges and often gathering them into a list is the process of edge tracing. A constant direction, clockwise or counter-clockwise, is used to perform this around the items. One of the edge tracing approaches is chain coding (Freeman, 1974). If you want to calculate shape measurements or identify or categorize an item, you may utilize the non-raster representation that comes out of this process. The approach is affected by the object's rotation and size, however.

Our description of the object in this work is based on the turning angle representation. The shape representation is shown in Figure 6 as a collection of turning angles □ = {□ (1), □ (2), □ (3),..., □ (N)}. This technique remains unchanged regardless of how the item is sized or moved. Furthermore, upon normalization, it remains unchanged regardless of rotation. A delicate edge's expression may be conveyed by the object's turning angle.

variations, such as those in length and radius. Classifying feature tokens according to their curvature qualities, we take turning angle variation into account. The initial step in our technique is to divide the edge into a certain number of segments. Computing the local maximum curvature points in each segment allows us to identify the turning sites from the edge. Consequently, there is a constant amount of picture turning points. Instead of picking points at the segment points, when crucial turning angles could be lost, our technique is superior. Normalizing the shape is a prerequisite before doing similarity measurement. The lengths of the edge segments are first normalized. Next, we locate the object's centroid, also known as its center of mass. The picture in Figure 7 is worth thinking about. Following this, a weight graph representing the vertical profile (an orthogonal projection on the X-axis) will be shown. The mass's mid-line may be simply calculated. On the left side of Figure 7 is the horizontal profile, which is an orthogonal projection on the Y-axis. The object's centroid can be easily located. Finding the boundary point that is farthest from the centroid—and thus the beginning point of the turning angle representation—follows the centroid's definition (see Fig. 8).

1. The Implemented System

Our dynamic link library (DLL) and query interface were built using Microsoft Visual Basic and C++. Built on top of Microsoft SQL Server, our image database integrates with the query interface via ODBC (Open Database Connectivity).

As shown in Figure 11, the client-side query interface and an example query about the spatial relations are shown. After sketching a triangle, a square, and a circle using the drawing tools, the user may tweak the feature weights and the spatial relations among these three objects in the query area. After the user finishes the query picture, they may access the candidate photos and similarity scores in the candidate section at the bottom of the window by clicking the query button. Clicking on the candidate picture will expand it in the top right corner of the window, allowing the viewer to examine image details.

This example just demonstrates how a content-based system makes advantage of geographical relations. When dealing with information that is not yet known, the qualitative method offers more leeway. In other words, it is fairly uncommon for the user to have doubts about the extracted image material. An strategy that may be used to convey such an unclear inquiry is the qualitative approach. Our CBR system excels in that area.

IV. Conclusion

A fresh method for drawing pictures out of color picture databases using information about shape, location, and color has been suggested. In order to enhance the effectiveness of color-sensation and chromatology-based picture retrieval, we have presented a color-clustering method and an image database hierarchy. Additionally, we have updated the Seed Filling algorithm to better extract spatial and geometric information from images for similarity measurement. Because measuring distances between items is challenging, we instead rely on qualitative spatial interactions to identify commonalities between them. On top of that, there is a visual interface that is easy to use and a learning system that provides feedback. The system achieves good results while retrieving information from the picture database over the Internet after training to change attribute weights. Windows 98 is the platform on which the system has been installed. We examined five thousand images. As anticipated, the accuracy and performance were satisfactory. While this study has not focused on the shape acquiring approach, there are alternative ways that may be used to enhance it. We want to add several form retrieval algorithms to the system in future studies and provide users the freedom to choose the one that works best for their specific area of application.

REFERENCES

With contributions from E. M. Arkin, L. Chew, D. Huttenlocher, K. Kedem, and J. Mitchell (1991) One measure for comparing polygonal forms that can be efficiently computed. Volume 13, Issue 3, pages 209–216, of the IEEE Transactions on Pattern Analysis and Machine Intelligence.

Termes, P., Kreiss, J., and Alshuth, M. (1996) Image and video retrieval using IRIS. Amsterdam, Netherlands: First International Workshop on Image Database and Multimedia Search, pp. 170-179.

Tobias Bimbo (1999) Information Visualization. The publication was made by Academic Press in San Francisco, California, USA.

Hierarchical chamfer matching: a parametric edge-matching technique was published by Borgefors in 1988. Volume 10, Issue 11, Pages 849–865, of the IEEE Transactions on Pattern Analysis and Machine Intelligence.

Two novel methods for image matching: parametric correspondence and chamfer matching were introduced by Barrow, H. G. (1977). Massachusetts, USA, for the Fifth Annual International Joint Conference on Artificial Intelligence.

J. R. Bach, A. Gupta, B. Horowitz, R. Humphrey, A. Hampapur, C. Fuller, and R. Jain (1996) Virage, an open-source picture management framework, is an image search engine. Proceedings of the SPIE Conference on Storage and Retrieval of Still Images and Videos, 2760, 76–87.

A computational method for edge detection (Canny, 1986). PAMI-8(6), 679-698, published in the IEEE Transactions on Pattern Analysis and Machine Intelligence.

W. C. H. Chuang and C. C. J. Kuo (1996) Theory and applications of wavelet descriptors for planar curves. 56–70. Published in the January issue of the IEEE Transactions on Image Processing.

Dowe, J. (1993) Multimedia image retrieval based on content. Location: Boston, Massachusetts, USA, SPIE Conference on Visual Communication and Image Processing... J. Hafner, D. Lee, D. Petkovic, D. Steele, M. Gorkani, Q. Huang, B. Dom, J. Ashley, M. Flicker, H. Sawhney, W. Niblack, J. (1995) Reference: IEEE Computer Magazine, 28(9), 23-32. QBIC system: querying based on video and picture material.

In Computer Survey, 6, 57–97, Freeman (1974) discusses the processing of line drawing pictures by computers. Calculating invariants by means of elimination techniques was published in 1995 by Kapur, D., Y. N. Lakshman, and T. Saxena. Int'l. Conf. on Image Proceedings, Washington, D.C., USA.

"Properties of temporal constraint network for supporting distributed multimedia applications" was published in 1999 by Keh, H. C., T. K. Shih, and A. Y. Chang. Information Management and Computer Science, 2, 15-26. Netra: a toolset for browsing huge picture datasets (1997) by Ma, W. Y. and B. S. Manjunath. ICIP'97, pages 568–571, Santa Barbara, California, USA.

J. Shepherd, A. H. H. Ngu, and M. Nabil (1996) Used 2D projection interval representation to retrieve images with similarity. Information and Data Engineering (IEEE) Transactions, vol. 8, no. 4, pp. 533–539.

In 1997, X. Li, M. T. Ozsu, and Y. Niu proposed 2D-h trees as an index strategy for multimedia content-based picture retrieval. Pages 1710–1715 of the IEEE International Conference on Intelligent Processing Systems, Beijing, P.R.C.

Pei, S. C., and Shiue, H. L. (1998) Color image database indexing and retrieval. Taipei, Taiwan, R.O.C., Conference on Computer Vision, Graphics and Image Processing, pp. 196-199.

"Photobook: content-based" was published in 1996 by Pentland, Picard, and Sclaroff. processing of picture archives. Vol. 18, No. 3, pp. 233-254, International Journal of Computer Vision. The use of Fourier descriptors for shape identification was proposed by Person and K. S. Fu in 1977.

Proceedings of the IEEE Transactions on Systems and Man, 7, 170-179. One best linear operator for detecting step edges was proposed by J. Shen and S. Castan in 1992. Volume 54, Issue 2, Pages 112-133, Journal of Computer Vision, Graphics, and Image Processing—Graphi-cal Model and Understanding.

Interval algebras in space and time (Shih, T. K. and A. Y. Chang, 1998). Tokyo, Japan, site of the 12th International Conference on Information Networking. A completely automated content-based picture query system, VisualSEEK was developed by Smith, J. R., and S. F. Chang in 1996. Proceedings of the ACM International Conference on Multimedia, Boston, MA, USA, pp. 87-98.

Color index was developed by M. J. Swain and H. Ballard in 1991. The article is published in the Internal Journal of Computer Vision and measures 7(1) pages 11-32. Y. Theodoridis, T. Sellis, and M. Vazirgiannis (1996) geographic and chronological arrangement for use in multimedia programs. Workshop on International