Graph/Image Legend Retrieval†

YING-WEN HUANG[1], CHEN-CHUNG LIU[2], AND SHU-YUAN CHEN[1,*]
[1]*Department of Computer Science and Engineering, Yuan Ze University, Taiwan*
[2]*Graduate Institute of Network Learning Technology, National Central University, Taiwan*

ABSTRACT

There are many content-based retrieval methods for image databases, however, none of them have coped with both graph and image simultaneously. Moreover, existing graph retrieval methods handle either a silhouette or a graph component rather than the whole graph. Hence, it is the goal of this paper to propose a graph/image legend retrieval method. The proposed method consists of two phases: legend extraction and legend retrieval. In the extraction stage the legend images are first converted from RGB into YIQ color spaces to get *Y* values as gray-level images. The foreground is then separated from background for the legend image so as to determine the characteristics of the legend for later extraction and retrieval. Since each legend may not occupy the whole legend image, the region enclosing the legend only should be detected first to increase retrieval accuracy. In the retrieval stage, features are extracted for each legend to proceed legend matching. The features used in our method include aspect ratio, number of legend components and spatial histograms of pixel number, border length and gray level. Since the processed legends can be properly divided into two categories: graph and image, type-based matching is adopted to evaluate the similarity between the query legend and each database legend using different similarity measures according to the type of the query legend which can be automatically determined. In this way, the correct legend can be retrieved. The effectiveness and practicability of the proposed method have been demonstrated by various experiments.

Key words: legend retrieval, graph retrieval, image retrieval, content-based image retrieval, legend extraction, type-based matching.

1. INTRODUCTION

With the advent of computing technology, media acquisition and storage devices, and multimedia compression standards, more and more digital data are generated and available to users all over the world. Nowadays, it is easy to access electronic books, electronic journals, web portals, and video streams. Hence, it will be convenient for users to provide related information retrieval for the query legend. Unfortunately, although there are many content-based retrieval methods for image databases, none of them is specifically designed for coping with graph and image simultaneously, and handling the whole legend rather than a silhouette or a graph component.

There are many content-based retrieval approaches for image databases (Muneesawang & Guan, 2004; Dong & Bhanu, 2005; Han, Ngan & Zhang, 2005). In general, they can be divided into shape-based (Petrou & Kadyrov, 2004; Bartolini, 2005; Chalechale, Naghdy & Mertins, 2005; Erol & Kossentini, 2005), image-based (Swain & Ballard, 1991; Nezamabadi-pour & Kabir, 2004; Bishnu,

† This work was partially supported by the National Science Council of Taiwan, R.O.C., under Grants NSC 95-2745-E-155-008-URD.

* Corresponding author. E-mail: sychen@aa.nctu.edu.tw

Bhattacharya, Kundu, Murthy & Acharya, 2005) and region-based approaches (Jing, Li, Zhang & Zhang, 2004; Prasad, Biswas & Gupta, 2004). The shape-based method is concerned with a silhouette rather than our method for legend, which is comprised of a lot of curves or lines. The image-based method usually adopts similarity measures based on a color histogram. However, the legend may be colored or monochrome so a gray-level histogram should be used, even though this is not as powerful as a color histogram. The region-based approach should segment the query and database images into semantic regions that would be nonsensical for a graph.

Some approaches (Fletcher & Kasturi, 1988; Boatto et al., 1992; Joseph & Pridmore, 1992; Vaxiviere & Tombre, 1992; Lai & Kasturi, 1994) focus on managing the graphs that exist in documents, such as company logos, engineering diagrams, maps, business charts, fingerprints, musical scores, and so on. Interpretation of land register maps is proposed in Boatto et al. (1992). The problem of separating a text string from mixed text/graphics images is resolved in Fletcher and Kasturi (1988). Mechanical engineering drawings are discussed in Joseph and Pridmore (1992), Vaxiviere and Tombre (1992), Lai and Kasturi(1994). Those approaches are concerned with the topics related to the analysis and recognition of graphic components rather than the whole graph retrieval as proposed by our method.

It is the goal of this paper to propose a legend retrieval for handling graph and image retrieval simultaneously. Hence, the main difference between the proposed legend retrieval and other existing image retrieval methods are as follows. First, our approach separates the foreground from the background and identifies the legend part for matching so as to increase retrieval accuracy. Second, type-based matching is proposed in this study to handle graph and image retrieval simultaneously rather than coping with only graphs (Fletcher & Kasturi, 1988; Boatto et al., 1992; Joseph & Pridmore, 1992; Vaxiviere & Tombre, 1992; Lai & Kasturi, 1994) or images (Swain & Ballard, 1991; Hadjidemetriou, Grossberg & Nayar, 2004; Jing et al. 2004; Li & Wang, 2004; Liapis & Tziritas, 2004; Lu & Lim, 2004; Muneesawang & Guan, 2004; Naqa, Galatsanos & Wernick, 2004; Nezamabadi-pour & Kabir, 2004; Petrou & Kadyrov, 2004; Prasad et al., 2004; Bartolini, 2005; Bishnu et al., 2005; Chalechale et al. 2005; Daschiel & Datcu, 2005; Dong & Bhanu, 2005; Erol & Kossentini, 2005; Han et al. 2005; Vasconcelos & Lippman, 2005) individually, as in other retrieval systems. Third, the proposed method can handle whole graphs including many curves and lines rather than only a silhouette (Petrou & Kadyrov, 2004; Bartolini, 2005; Chalechale et al. 2005; Erol & Kossentini, 2005) or a graph component, as in other methods (Fletcher & Kasturi, 1988; Boatto et al., 1992; Joseph & Pridmore, 1992; Vaxiviere & Tombre, 1992; Lai & Kasturi, 1994).

The processed legends can be properly divided into two categories: graph and image. The proposed method consists of two phases: legend extraction and legend retrieval. In the extraction stage, the legend images are first converted from RGB into YIQ color spaces to get Y values as gray-level images. The foreground is then separated from the background for the legend image so as to get the characteristics of the legend for later legend extraction and retrieval. Since each legend may not occupy the whole legend image, the region enclosing the legend only should be detected first to increase retrieval accuracy. In the retrieval stage, features are extracted for each legend to proceed legend matching. The features used in our method include aspect ratio, number of legend components and spatial histograms of pixel number, border length and gray level. Finally, the strategy of type-based matching is adopted to evaluate the similarity between the query legend and each database legend using different similarity measures according to the type of query legend. In this way, the correct legend can be retrieved, which in turn facilitates the information retrieval for the legend. The flowchart of the proposed method is shown in Figure 1.

In Section 2, a legend database is introduced then, in Sections 3 and 4, we describe the proposed legend extraction and legend retrieval methods. In Section 5, we show the experimental results and finally, in Section 6, conclusions and future work are elucidated.

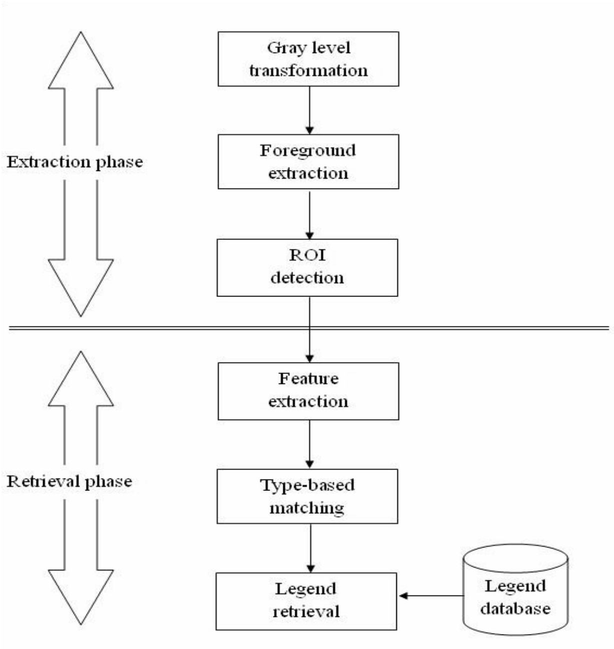
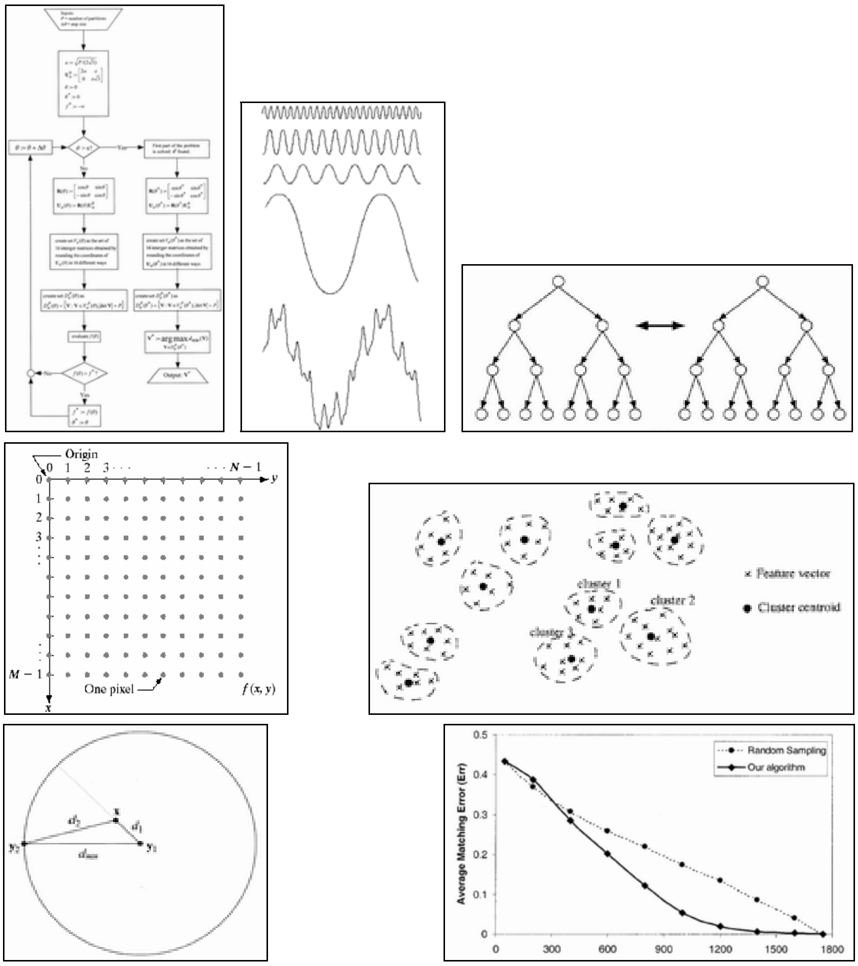


Figure 1. The flowchart of our method.

2. LEGEND DATABASE

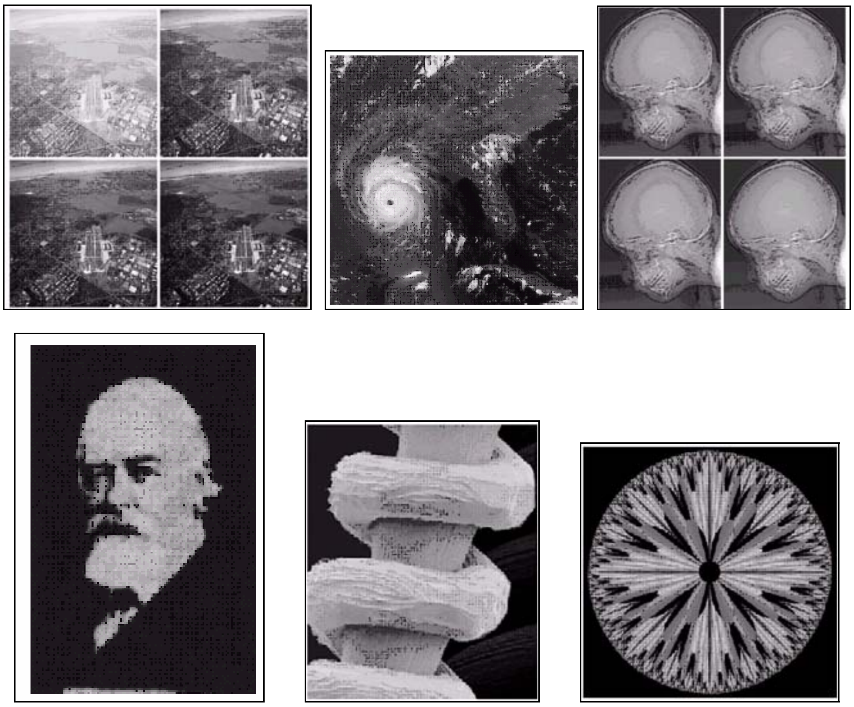
The processing unit of the proposed method is a legend which may exist in electronic books, video mediums, or web portals. The processed legends can be properly divided into two categories: graph and image. In general, graph-legends consist of simple lines and curve, while the image-legends may be of diverse form. Some examples of legends are shown in Figure 2.

In this study, there are no limitations for the size of legend image, which can be variable. Legend images were collected from Power-Point files or electronic books to construct a database for experiment. The number of images in our database is 450 including 250 graph and 200 image legends. The legend existing in each image of database is also scanned from books as the query image.



(a) Graph legend.

Figure 2. Some examples of legends. (Courtesy of Gonzalez and Woods (2002))



(b) Image legends.

Figure 2 (continued).

3. LEGEND EXTRACTION

3.1 Gray Level Transformation

Since the legend may be colored or monochrome, the legend images should be transformed into the standard form so that a comparison between the two kinds of legends is possible. In this study, colored images are transformed into gray-level images by transferring RGB color space into YIQ color space. The transformation can be explicitly described by the following equation:

$$\begin{bmatrix} Y \\ I \\ Q \end{bmatrix} = \begin{bmatrix} 0.299 & 0.587 & 0.114 \\ 0.596 & -0.275 & -0.321 \\ 0.212 & -0.528 & 0.311 \end{bmatrix} \begin{bmatrix} R \\ G \\ B \end{bmatrix}. \tag{1}$$

Henceforth, we can get a gray-level legend image, g(x, y), with Y value as the gray value.

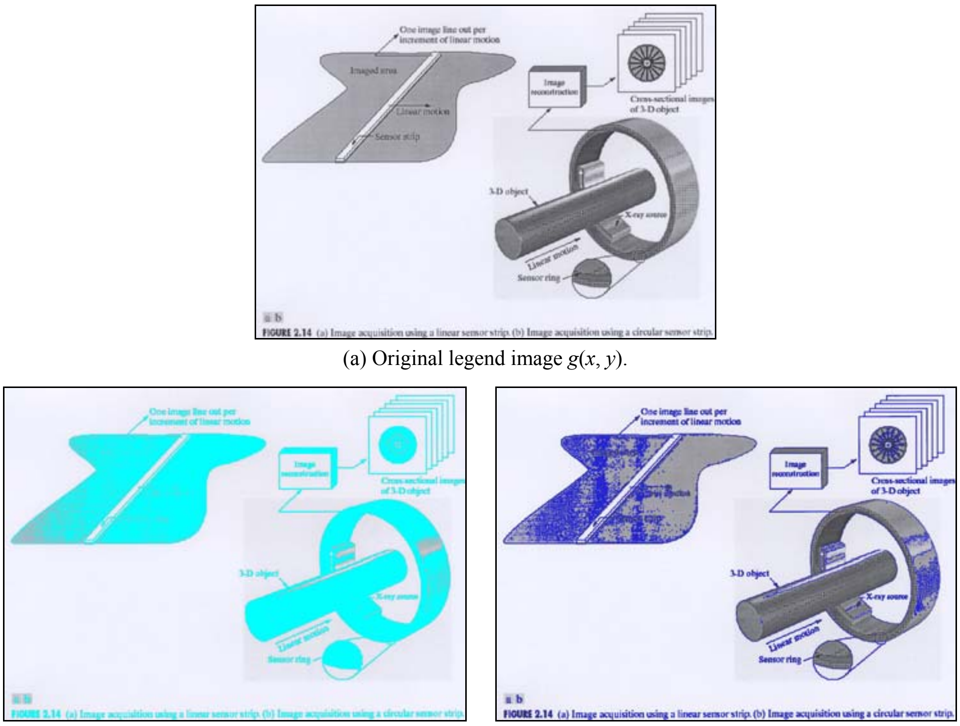
3.2 Foreground Extraction

For legend retrieval, the foreground of a legend image should be extracted first. In this study the foreground is specified by a foreground map and foreground border images. Simple bi-level thresholding and border detector are adopted to acquire them. More specifically, the foreground map and border images, *f*(*x*, *y*) and *b*(*x*, *y*) can be defined by the following equations:

$$f(x, y) = \begin{cases} 1, & g(x, y) \le TH_b; \\ 0, & \text{otherwise}; \end{cases} \tag{2}$$

$$b(x, y) = \begin{cases} 1, & g(x, y)\ \text{satisfys border criterion}\ B(x, y); \\ 0, & \text{otherwise}; \end{cases} \tag{3}$$

where TH_b is determined automatically by the proposed method as described in Section 3.2.1 and *B*(*x*, *y*) is the criterion of border points as described in Section 3.2.2. Examples of foreground map and border images are shown in Figure 3.


One image line out per increment of linear motion
Imaged area
Linear motion
Sensor strip
Image reconstruction
Cross-sectional images of 3-D object
3-D object
X-ray source
Linear motion
Sensor ring
a b
FIGURE 2.14 (a) Image acquisition using a linear sensor strip. (b) Image acquisition using a circular sensor strip.

(a) Original legend image *g*(*x*, *y*).

(b) Foreground map image *f*(*x*, *y*).

(c) Foreground border image *b*(*x*, *y*).

Figure 3. An example of foreground map and border images.

3.2.1 Foreground map image

The essential factor for bi-level thresholding is threshold determination. The threshold value is crucial to separate foreground from background, which in turn dominates the retrieval accuracy. The difficulty in determining the threshold is due to the following factors. First, the images of legends may be captured under different illuminations and environments. Second, different resources may have different signal magnitudes. In this study, moment-preserving thresholding (Tsai, 1985) is adopted and modified to handle these problems so as to obtain a good foreground map.

The moment-preserving thresholding (Tsai, 1985) is first briefly described followed by its modified version. Given an image *g*(*x*, *y*) with *n* pixels, the *i*th moment m_i of *g*(*x*, *y*) is defined as:

$$m_i = \left(\frac{1}{n}\right)\sum_x\sum_y \left(g(x,y)\right)^i,\ i = 1, 2, 3, \cdots \tag{4}$$

Let *g*(*x*, *y*) with values less than and greater or equal to TH_b be replaced by z_0 and z_1, respectively in the bi-level image. Let the fractions of them be p_0 and p_1, respectively. The threshold value TH_b can then be chosen as p_0-tile of the histogram of *g*(*x*, *y*), where p_0 can be calculated by:

$$p_0 = \left(\frac{1}{p_d}\right)\begin{vmatrix} 1 & 1 \\ m_1 & z_1 \end{vmatrix}, \tag{5}$$

$$p_d = \begin{vmatrix} 1 & 1 \\ z_0 & z_1 \end{vmatrix},\ z_1 = \left(\frac{1}{2}\right)\left[-c_1 + \sqrt{\left(c_1^2 - 4c_0\right)}\right],$$

$$c_0 = \left(\frac{1}{c_d}\right)\begin{vmatrix} -m_2 & m_1 \\ -m_3 & m_2 \end{vmatrix},\ c_1 = \left(\frac{1}{c_d}\right)\begin{vmatrix} m_0 & -m_2 \\ m_1 & -m_3 \end{vmatrix},$$

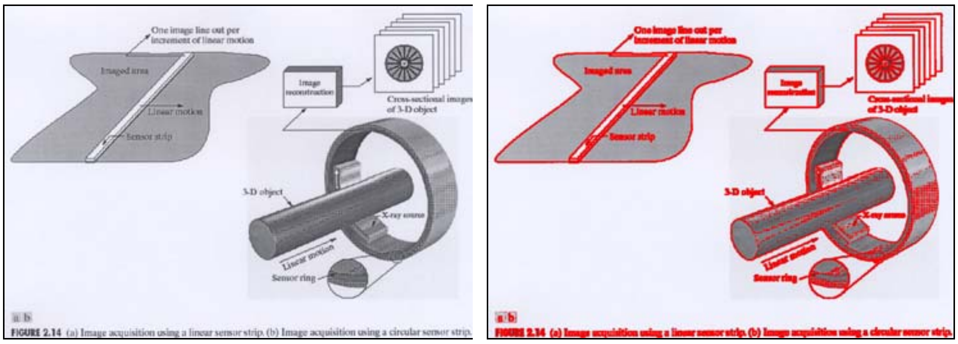
$$c_d = \begin{vmatrix} m_0 & m_1 \\ m_1 & m_2 \end{vmatrix}.$$

The details of derivation can be referred to Tsai (1985).

In our method we further utilize a Sobel detector (Gonzalez & Woods, 2002) to find the proper threshold values for legend retrieval. Assume that resulting image of applying the Sobel detector to the legend image *g*(*x*, *y*) is *s*(*x*, *y*). We can then get edge image *e*(*x*, *y*) as described by:

$$e(x,y) = \begin{cases} 1, & s(x,y) \geq TH_e; \\ 0, & \text{otherwise.} \end{cases} \tag{6}$$

In this study, TH_e is set to 127. An example of edge images is shown in Figure 4.


One image line out per increment of linear motion
Imaged area
Linear motion
Sensor strip
Image reconstruction
Cross-sectional images of 3-D object
3-D object
X-ray source
Linear motion
Sensor ring
a b
FIGURE 2.14 (a) Image acquisition using a linear sensor strip. (b) Image acquisition using a circular sensor strip.

(a) Original legend image $g(x, y)$.

(b) The resulting edge image $e(x, y)$.

Figure 4. An example of edge images.

We then re-execute moment preserving thresholding on the image $h(x, y)$

$$h(x,y)=\begin{cases} g(x,y), & e(x,y)=1; \\ 0, & \text{otherwise.} \end{cases} \tag{7}$$

In other words, only the gray levels of edge pixels are of concern for threshold determination. The thresholding results on $g(x, y)$ and $h(x, y)$ are shown in Figure 5. From this figure, we found that the thresholding results on $h(x, y)$ are better than those on $g(x, y)$.

3.2.2 Foreground border image

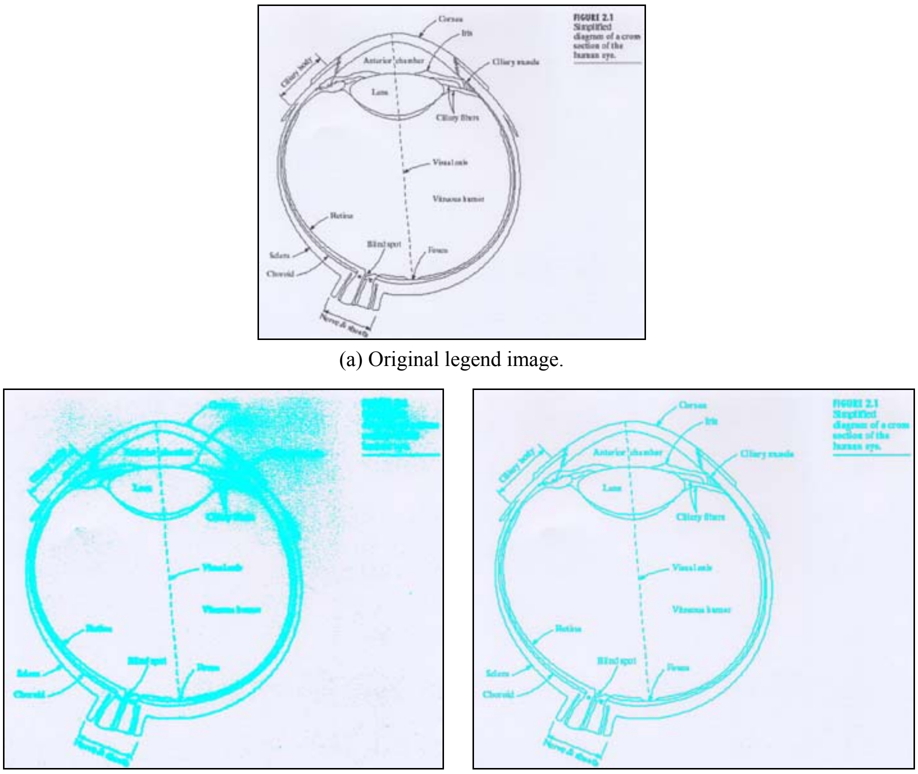
In this study, border detector (Gonzalez & Woods, 2002) is applied to the foreground map image $f(x, y)$ to get the border image $b(x, y)$. The border detector is performed by scanning $f(x, y)$ from top to bottom and from left to right. More specifically, for each pixel $f(x, y)$, if one of the upper neighbors $f(x, y\text{-}1)$ or left neighbors $f(x\text{-}1, y)$ has different binary labels from $f(x, y)$, the point (x, y) is denoted as a border point. Thus, the border criterion $B(x, y)$ can be specified by:

$$B(x,y)=\left(f(x,y)\oplus f(x-1,y)\right)\vee\left(f(x,y)\oplus f(x,y-1)\right) \tag{8}$$

where $\oplus$ and $\vee$ denote the exclusive or and or operators, respectively. An example of border image is shown in Figure 3(c).

3.3 Region of Interest

Sine each legend may not occupy the whole legend image, the region enclosing the legend only should be detected first to increase retrieval accuracy. In other words, it is necessary to find the region of interest (ROI) according to the boundary of the legend. Projection technology is adopted in this study to solve this problem. The ROI detection consists of two stages: content and legend detection to find a region of content (ROC) and a region of legend (ROL).



(a) Original legend image.

(b)Results of thresholding on $g(x, y)$.

(c) Results of thresholding on $h(x, y)$.

Figure 5. Comparison of thresholding results on $g(x, y)$ and $h(x, y)$.

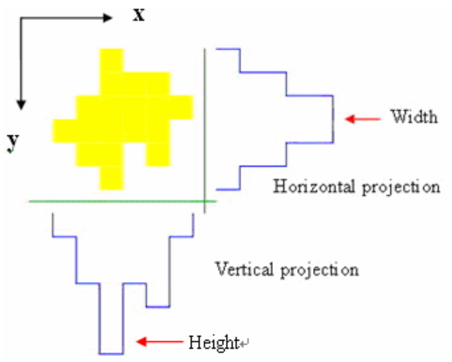


Figure 6. Example of horizontal and vertical projections.

3.3.1 Content detection

The projection in horizontal and vertical directions can be used to find the horizontal and vertical boundaries. Since the methods for both directions are similar, we describe the method for the horizontal case only. The horizontal and vertical projections $p_v(x)$ and $p_h(y)$ are defined by:

$$p_v(x) = \sum_y f(x, y);$$
$$p_h(y) = \sum_x f(x, y). \tag{9}$$

Let the legend image have a width of W and a height of H. The most top and bottom rows having horizontal projections $p_h(y)$ greater than a threshold value TH_p are regarded as the horizontal boundary of the ROC. In this study, TH_p is set to 3. The same method can be used in the vertical detection. In this way, we can find the boundary of the ROC.

3.3.2 Legend detection

Since the ROC may or may not include a caption, the caption part should be detected and deleted so that the matching evaluation can be performed on only the legend part (ROL) to increase retrieval accuracy. The caption part can also be detected by the projection technology.

Let the ROC superimpose on the foreground map image $f(x, y)$. The horizontal projections $p_h^r(y)$ or vertical projections $p_v^r(x)$ during the ROC are computed. The first and the second rows having $p_h^r(y)$ greater than a threshold value TH_p are regarded as the horizontal border of the first candidate, and the third and the fourth as the second candidate and so on. The vertical borders of candidates can be obtained in the similar way. Each candidate is then verified by the following constraints in sequence to determine whether it is a caption or legend.

(1) Legend constraint

The constraint is derived from the observation that a caption is comprised of text and text has short vertical lines as opposed to the long vertical lines of a graph. Those candidates with long vertical lines will not be a caption. The details of this constraint are described as follows.

Let the candidate have width of W_c and height H_c. Let each candidate superimpose on the foreground map image $f(x, y)$. The vertical projections $p_v^c(x)$ during the candidate are computed. If there is a column having $p_v^c(x)$ greater than a threshold value W_c /16 or a sum of $p_v^c(x)$ belonging to three consecutive columns greater than a threshold value W_c /14, the candidate is considered as a legend.

(2) Caption constraint

The constraint is derived from the observation that a caption is composed of text. Those candidates having too few or too many pixels in a foreground map image will not be a caption. The details of this constraint are described as follows.

Let each candidate superimpose on the foreground map image *f*(*x*, *y*). The number of pixels, $N_c^{(f)}$, covered by the candidate is then counted. The pixel number ratio related to the candidate, $R_c^{(f)}$, can be defined as:

$$R_c^{(f)} = \frac{N_c^{(f)}}{W_c \times H_c}. \tag{10}$$

Those candidates with a pixel number ratio, $R_c^{(f)}$, less than a threshold value TH_l or larger than a threshold value TH_u are regarded as captions and deleted. In this study, TH_l and TH_u are set to 0.058 and 0.3, respectively. After captions are deleted, the ROC detection as described in Section 3.3.1 is applied again on those remaining parts of the ROC to obtain the graph/image part, i.e., ROL. Examples of ROL detection are shown in Figures 7 and 8.

4. LEGEND RETRIEVAL

The type of query legend is first determined as one of the three types: graph, image and hybrid of graph and image according to the number of foreground pixels in the query legend. Different retrieval strategies will then be adopted according to the type of query legend.

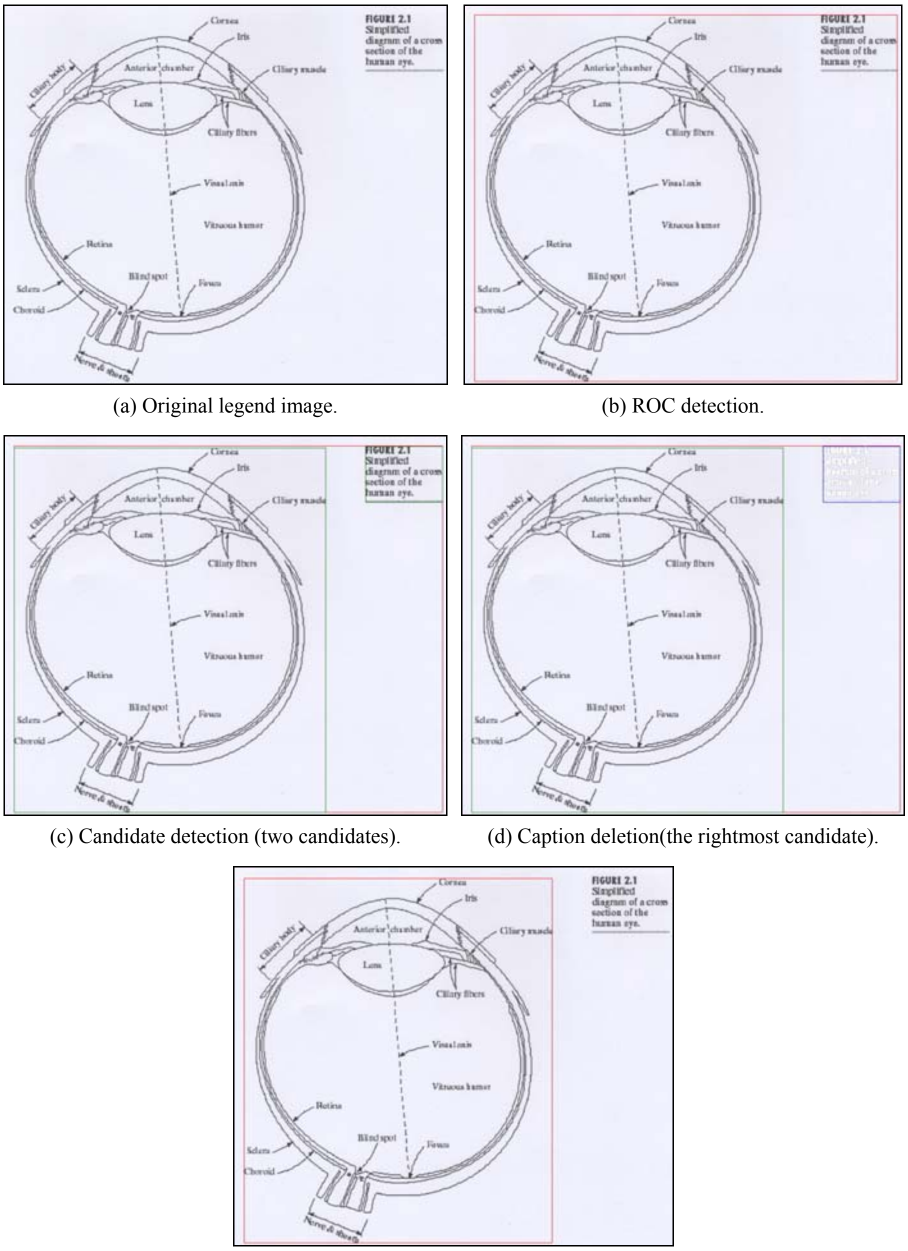
4.1 Feature Extraction

The features used in this study for legend retrieval include aspect ratio, number of legend components and spatial histograms, derived from a foreground map, border and quantized images, namely histograms of pixel number, border length and gray level, respectively. They are defined as follows:

Let the ROL have a width of W_l and a height of H_l. The aspect ratio of ROL, *Asp*, can be defined as:

$$Asp = \frac{W_l}{H_l}. \tag{11}$$

Let ROL superimpose on the foreground map image *f*(*x*, *y*). The horizontal projections $p_h^l(y)$ or vertical projections $p_v^l(x)$ during the ROL are computed. The first and the second rows having $p_h^l(y)$ greater than a threshold value TH_p are regarded as the horizontal border of a legend component, and the third and the fourth as the second components, and so on. For each horizontal legend component, the vertical borders of each component can be obtained in the similar way. The number of legend components is then counted as a feature *Noc*. Examples of *Noc* are shown in Figure 9.



(a) Original legend image.

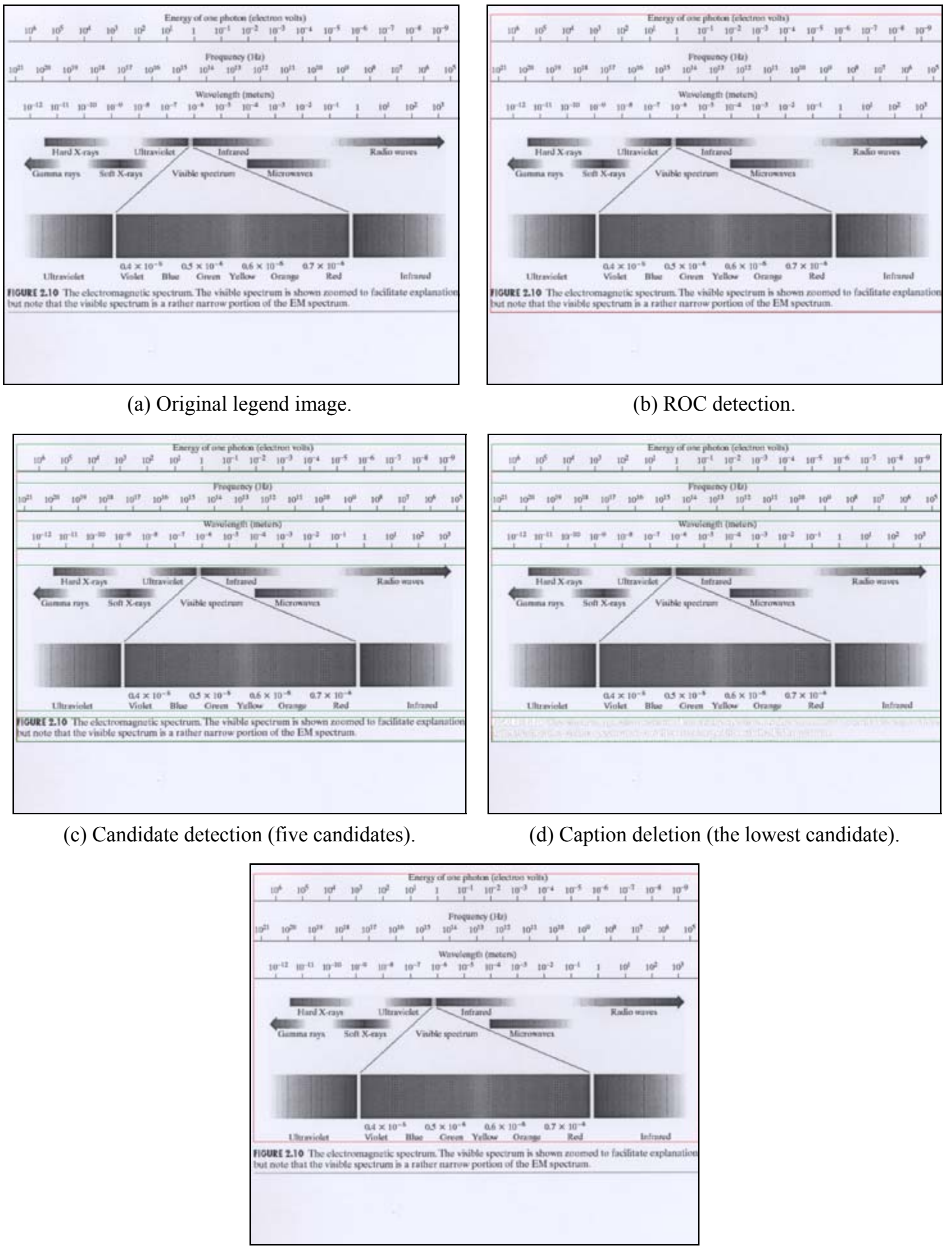
(b) ROC detection.

(c) Candidate detection (two candidates).

(d) Caption deletion(the rightmost candidate).

(e) ROC detection again to obtain ROL.

Figure 7. An example of ROL detection.


Energy of one photon (electron volts)
Frequency (Hz)
Wavelength (meters)
Hard X-rays
Ultraviolet
Infrared
Radio waves
Gamma rays
Soft X-rays
Visible spectrum
Microwaves
0.4 × 10⁻⁶
0.5 × 10⁻⁶
0.6 × 10⁻⁶
0.7 × 10⁻⁶
Ultraviolet
Violet
Blue
Green
Yellow
Orange
Red
Infrared
FIGURE 2.10 The electromagnetic spectrum. The visible spectrum is shown zoomed to facilitate explanation but note that the visible spectrum is a rather narrow portion of the EM spectrum.

(a) Original legend image.

(b) ROC detection.

(c) Candidate detection (five candidates).

(d) Caption deletion (the lowest candidate).

(e) ROC detection again to obtain ROL.

Figure 8. The other example of ROL detection.

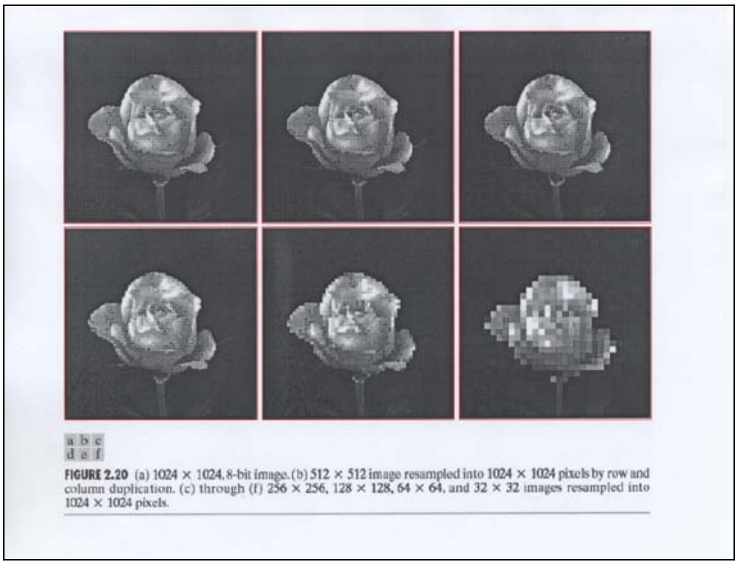


FIGURE 2.20 (a) 1024 × 1024, 8-bit image. (b) 512 × 512 image resampled into 1024 × 1024 pixels by row and column duplication. (c) through (f) 256 × 256, 128 × 128, 64 × 64, and 32 × 32 images resampled into 1024 × 1024 pixels.

(a) In this case, *Noc* = 6.

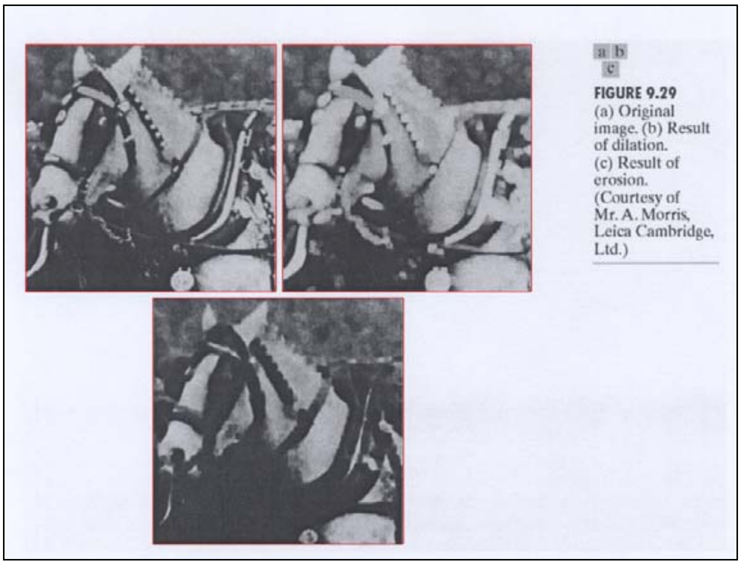


FIGURE 9.29
(a) Original image. (b) Result of dilation.
(c) Result of erosion.
(Courtesy of Mr. A. Morris, Leica Cambridge, Ltd.)

(b) In this case, *Noc* = 3.

Figure 9. Examples of *Noc*.

To obtain the histogram feature, the ROL is first partitioned into 4×4=16 blocks as shown in Figure 10. Let the 16 blocks be indexed in ascendant order from left to right and from top to bottom. The number of pixels in $f(x, y)$ enclosed by the *i*-th block is counted as $N_i^{(p)}$. The histogram of pixel number, $\mathbf{h}^{(p)}$, can be defined as:

$$\mathbf{h}^{(p)} = \left(h_i^{(p)}\right),$$
$$h_i^{(p)} = \frac{N_i^{(p)}}{16 \times W_l \times H_l},$$
$$i = 1,\ 2, \cdots, 16\ . \tag{12}$$

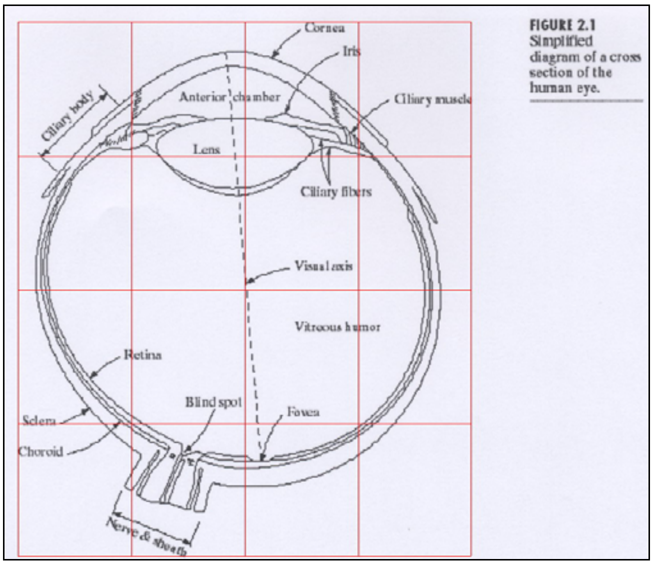


Figure 10. ROL is partitioned into 4×4 blocks.

Similarly, the histogram of border length, $\mathbf{h}^{(b)}$, can be calculated by the following equation except that the foreground map image *f*(*x*, *y*) is replaced by the foreground border image *b*(*x*, *y*):

$$\mathbf{h}^{(b)} = \left(h_i^{(b)}\right),$$
$$h_i^{(b)} = \frac{N_i^{(b)}}{16 \times W_l \times H_l}, \tag{13}$$
$$i = 1, 2, \cdots, 16\ ,$$

where $N_i^{(b)}$ is the number of border points in *b*(*x*, *y*) enclosed by the *i*-th block.

Finally, the histogram of the gray level is obtained as follows. First, the quantized image, *q*(*x*, *y*) of the input legend image *g*(*x*, *y*) should be obtained by the following quantization method. The maximum and minimum gray values, $g_{\max}$ and $g_{\min}$, in *g*(*x*, *y*) are computed. The gray values are then uniformly quantized into four levels by dividing the range of ($g_{\min}$, $g_{\max}$) into four equal intervals. More specifically, *q*(*x*, *y*) can be defined by:

$$q(x,y) = j, \text{ if } g(x,y) \in \left(g_{\min} + (j-1)\times\delta, g_{\min} + j\times\delta\right), j = 1, 2, 3, 4,$$
$$\delta = \frac{g_{\max} - g_{\min}}{4}. \tag{14}$$

The histogram of gray level, $\mathbf{h}^{(g)}$, is then easily obtained as follows. The number of pixels with quantized label *j* in *q*(*x*, *y*) enclosed by the *i*-th block is counted as $N_{i,j}^{(g)}$. The histogram of $\mathbf{h}^{(g)}$, can then be defined by:

$$\begin{aligned}
&\mathbf{h}^{(g)} = \left(h_{i,j}^{(g)}\right), \\
&h_{i,j}^{(g)} = \frac{N_{i,j}^{(g)}}{16\times W_l \times H_l}, \\
&i = 1, 2, \cdots, 16, j = 1, 2, \cdots, 4.
\end{aligned} \tag{15}$$

4.2 Type-based Matching

In the retrieval stage, the feature of aspect ratio, *Asp*, is first used to prune irrelevant database legends. The type-based matching can then be performed between the query and only the database legends in the small plausible set. More specifically, the similarity measure between query legend *Q* and database legend *S* on the basis of *Asp* is defined by:

$$d_{Asp}(Q,S) = \left|Asp(Q) - Asp(S)\right| \tag{16}$$

where (*Q*) refers to features of *Q* and (*S*) refers to features of *S*, respectively. Thus, we can get the plausible set *P*(*Q*) for *Q* as:

$$P(Q) = \left\{S \middle| d_{Asp}(Q,S) \le 0.2\right\}. \tag{17}$$

The type of *Q*, *T*(*Q*), can be determined according to the number of pixels by the following equation:

$$T(Q) = \begin{cases} \text{Graph}, & f(Q) < 0.25; \\ \text{Hybrid}, & 0.25 \le f(Q) \le 0.5; \\ \text{Image}, & 0.5 < f(Q), \end{cases} \tag{18}$$

$$f(Q)=\sum_{i=1}^{16}\frac{N_i^{(p)}(Q)}{W_l \times H_l}\ .$$

The final similarity measures, $d(Q, S)$, between Q and S are defined by

$$d(Q,S)=\begin{cases} d^{(p)}(Q,S)+d^{(b)}(Q,S), & T(Q)=\text{Graph;}\\ d^{(p)}(Q,S)+d^{(b)}(Q,S)+d^{(g)}(Q,S), & T(Q)=\text{Hybrid;}\\ d^{(g)}(Q,S)+d_{Noc}(Q,S), & T(Q)=\text{Image,}\end{cases}$$

$$d^{(p)}(Q,S)=\sum_{i=1}^{16}\left|h_i^{(p)}(Q)-h_i^{(p)}(S)\right|,$$

$$d^{(b)}(Q,S)=\sum_{i=1}^{16}\left|h_i^{(b)}(Q)-h_i^{(b)}(S)\right|, \tag{19}$$

$$d^{(g)}(Q,S)=\sum_{i=1}^{16}\sum_{j=1}^{4}\ \left|h_{i,j}^{(g)}(Q)-h_{i,j}^{(g)}(S)\right|,$$

$$d_{Noc}(Q,S)=\left|Noc(Q)-Noc(S)\right|.$$

Note that the feature *Noc* is defined in Section 4.1. Thus, the final similarity measure, $d(Q, S)$, between query legend Q and each database legend S in the set $P(Q)$ can be calculated by Eq. (19). It is noted that there are no weights in equation (19) since we found that different features have almost the same significance as the experimental results. Those database legends in the set $P(Q)$ are then sorted in ascending order of $d(Q, S)$. The top database legend in the list is regarded as the retrieval result for the query legend Q.

5. EXPERIMENTAL RESULTS

The proposed method has been implemented on a personal computer with a single AMD K-8 3200+ CPU and 512 Megabytes DDRAM. The operating system is Microsoft Windows XP Server Chinese version Service Pack 2. The program was developed in the C++ language and compiled under Borland C++ Builder version 6.

The performance of legend retrieval can be measured by the retrieval accuracy. The retrieval accuracy is computed as the ratio of the number of legends correctly retrieved to the number of total query legends. Moreover, not only the retrieval accuracy with respect to the first rank but also the second and third ranks are of concern in our experiments.

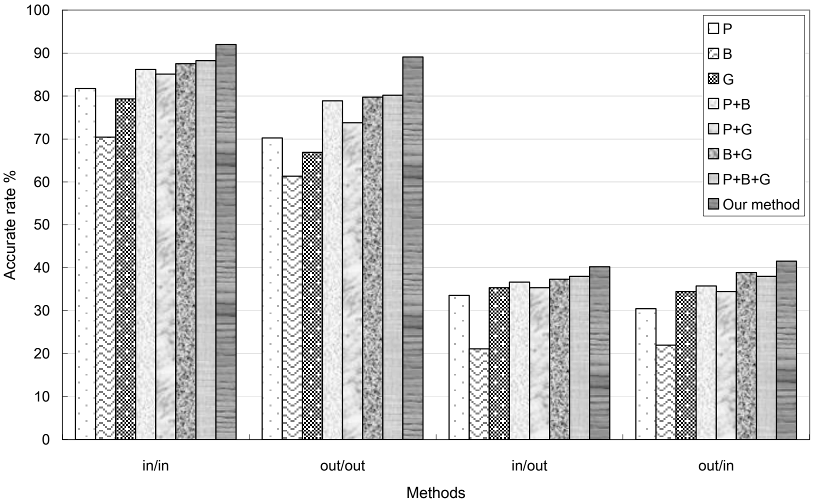


Figure 11. Retrieval accuracies in the cases of in/in, out/out, in/out, and out/in for the partition of 4×4 using different combinations (P, B, G, P+B, P+G, B+G, P+B+G and our method).

Some options need to be chosen in the proposed method, for example the partition size and the number of features. These will be determined by experimentation. In this study, the partitions of 3×3 and 4×4 will be used. The combinations of features include only pixel number, $\mathbf{h}^{(p)}$ (denoted P), only border length, $\mathbf{h}^{(b)}$ (denoted B), only gray level $\mathbf{h}^{(g)}$ (denoted G), and the different combinations of $\mathbf{h}^{(p)}$, $\mathbf{h}^{(b)}$, and $\mathbf{h}^{(g)}$ (denoted P+B, P+G, B+G, P+B+G). The necessity of type-based matching is also discussed. On the other hand, query and database legend images may include caption (denoted in) or may exclude caption (denoted ex). Thus, there will be four retrieval status: in/in, in/ex, ex/ex, ex/in and each *u*/*v* represents the query in the status of *u* and the database in the status of *v*.

First, the retrieval accuracies of 1st rank in the cases of in/in, out/out, in/out and out/in for the partition of 4×4 using different combinations of histograms (P, B, G, P+B, P+G, B+G, P+G+B) and our method are depicted in Figure 11. From this figure, we found that the combination of pixel size, border length, and gray levels, i.e., P+G+B out-performs the others except for our type-based matching method. In addition, we can conclude that the strategy of type-based matching is necessary since our method adopts this strategy and has the highest accuracy. Moreover, the accuracy of in/in is higher than those of ex/ex, in/ex, ex/in. It is not surprising since the query and database images in the case of in/in include more information than others.

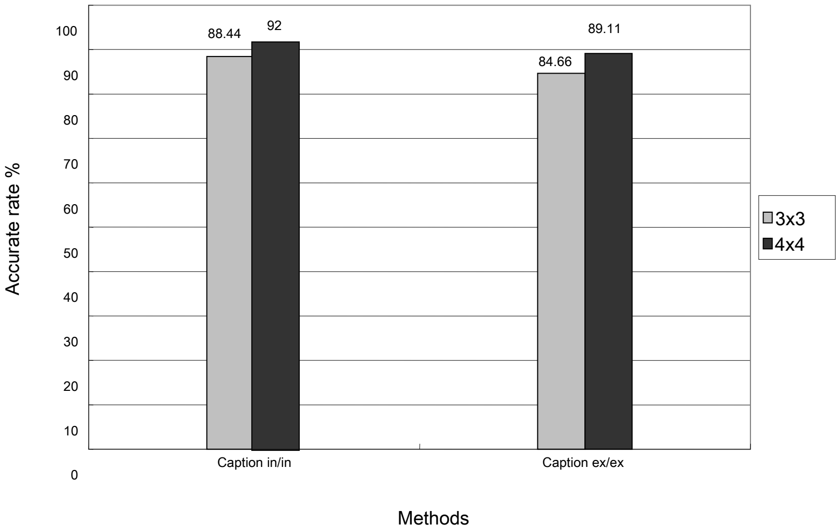


Figure 12. The retrieval accuracies of our method for the partitions 3×3 and 4×4.

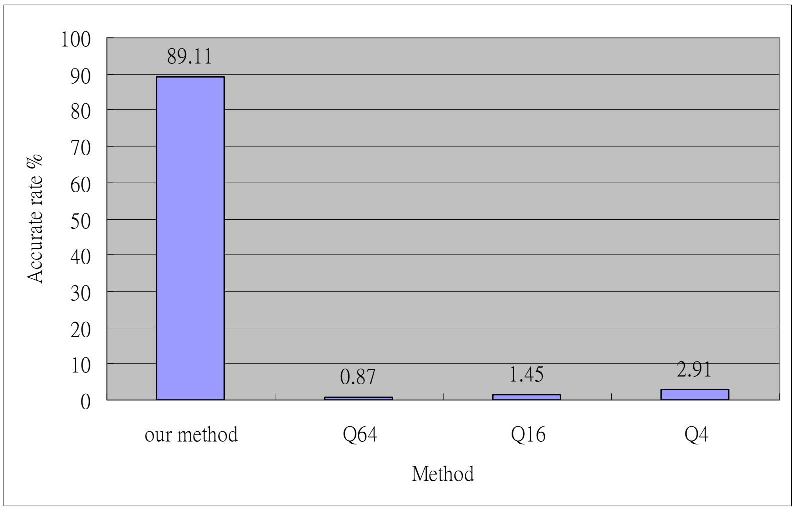
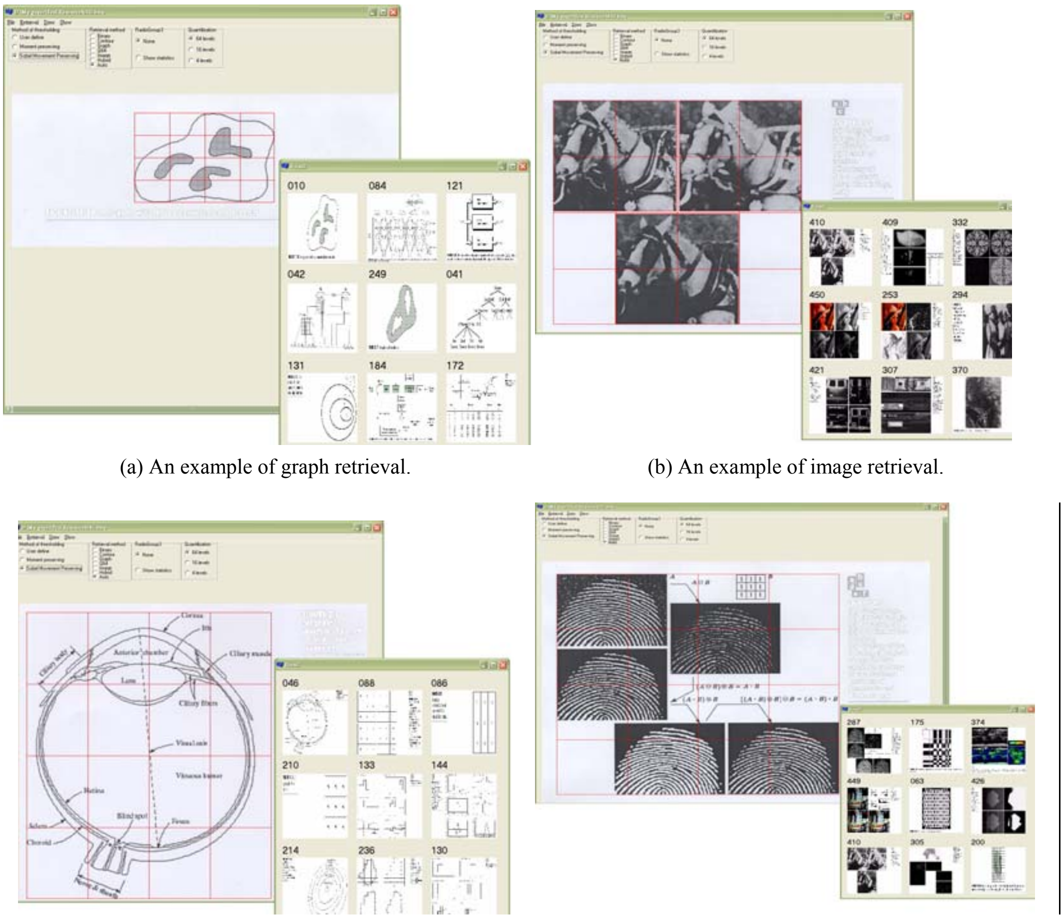


Figure 13. Comparison of our method to traditional image retrieval based on histogram matching (Swain & Ballard, 1991) (quantized levels of 64, 16, and 4, respectively).



(a) An example of graph retrieval.

(b) An example of image retrieval.

(c) An example of graph retrieval.

(d) An example of image retrieval.

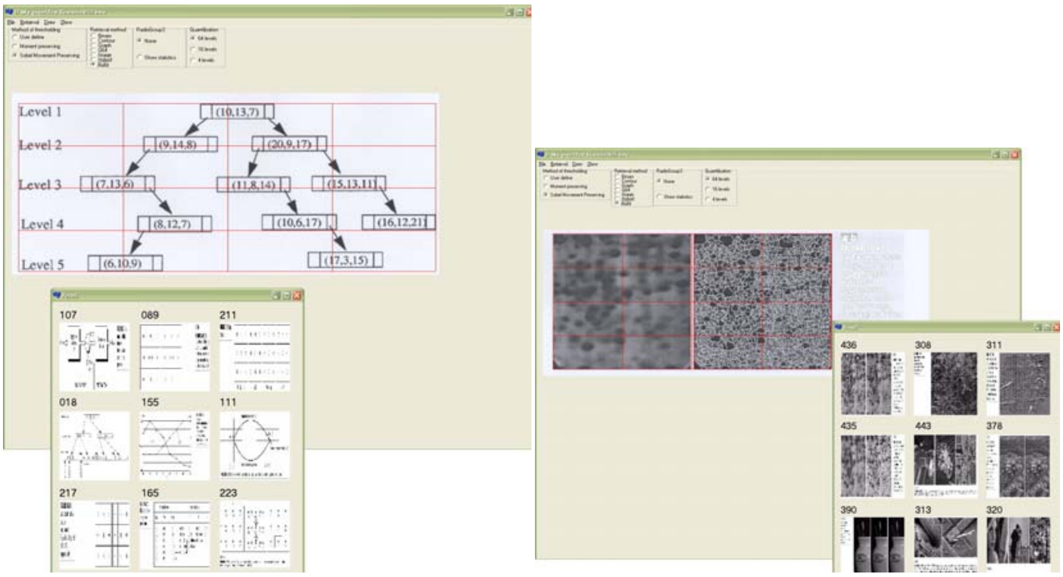
Figure 14. Examples of success retrieval results.

The retrieval accuracies of top 1 in the cases of in/in and ex/ex for the partitions of 3×3 and 4×4 using our method are depicted in Figure 12. From Figure 12, we found that the partition of 4×4 has higher accuracy rate. Thus the partition of 4×4 is chosen in the proposed method. Finally, the retrieval accuracies of our method with respect to top 1, top 2, and top 3 for the partition of 4×4 are listed in Table 1. From this table, we can conclude that the retrieval accuracy increase when more legends are retrieved.

The proposed method is compared to the traditional image retrieval based on histogram matching (Swain & Ballard, 1991) with resolutions of 64, 16, and 4 bins. The comparison results are shown in Figure 13. It is obvious that our method is far more accurate than the other methods. Some successful retrieval results are shown in Figure 14 and some failure cases are shown in Figure 15.

Table. 1. *Retrieval accuracies of our method in the cases of in/in, ex/ex, in/ex, and ex/in for the partition of 4×4*

		Top 1	Top 2	Top 3
in/in	Image	83.15	91.05	94.73
	Graph	98.46	98.84	99.61
	Total	92.00	95.55	97.55
ex/ex	Image	82.10	91.05	94.21
	Graph	94.23	96.15	96.92
	Total	89.11	94.00	95.77
in/ex	Image	40.52	49.47	53.68
	Graph	40.00	45.00	48.84
	Total	40.22	46.88	50.88
ex/in	Image	44.21	53.15	56.84
	Graph	39.61	43.07	46.15
	Total	41.55	47.33	50.66



(a) An example of graph retrieval.

(b) An example of image retrieval.

Figure 15. Examples of failure retrieval results.

6. CONCLUSIONS

A legend retrieval method is proposed in this paper to provide related information retrieval for the legend. The proposed method consists of two phases: legend extraction and legend retrieval. The former includes gray-level transformation, foreground extraction and legend extraction. The latter includes feature extraction and type-based matching. In this study, the adopted features include aspect ratio, number of legend components and spatial histograms of pixel number, border length and gray level. The main difference between the proposed legend retrieval and other existing image retrieval methods are as follows. First, our approach separates the foreground from the background and identifies the legend part for matching so as to increase retrieval accuracy. Second, type-based matching is proposed in this study to handle graph and image retrieval simultaneously as opposed to other systems that can cope with only graph or image individually. Third, the proposed method can handle whole graph including many curves and lines as opposed to other methods which can only handle a silhouette or a graph component. Experimental results demonstrate the effectiveness of the proposed method.

Future work can be directed to the following topics.

(1) More features should be extracted from the legend to improve the effectiveness of image-legend retrieval.

(2) More intelligent features should be proposed to discriminate the caption from the legend in the foreground.

(3) Extend the legend retrieval method for practical applications.

REFERENCES

Bartolini, I. (2005). Warp: accurate retrieval of shapes using phase of Fourier descriptors and time warping distance. *IEEE Trans. Pattern Analysis and Machine Intelligence*, *27*(1), 142-147.

Bishnu, A., Bhattacharya, B. B., Kundu, M. K., Murthy, C. A., & Acharya, T. (2005). Euler vector for search and retrieval of gray-tone image. *IEEE Trans. Systems, Man and Cybernetics*, *35*(4), 801-812.

Boatto, L., Consorti, V., Buono, M. D., Zenzo, S. D., Eramo, V., Esposito, A. Melcarne, F., Meucci, M., Morelli, A., Mosciatti, M., Scarci, S., & Tucci, M. (1992). An interpretation system for land register maps. *IEEE Trans. Computer*, *25*(7), 25-33.

Chalechale, A., Naghdy, G., & Mertins, A. (2005). Sketch-based image matching using angular partitioning. *IEEE Trans. Systems, Man and Cybernetics*, *35*(1), 28-41.

Daschiel, H., & Datcu, M. (2005). Information mining in remote sensing image archives: system evaluation. *IEEE Trans. Geoscience and Remote Sensing*, *43*(1), 188-199.

Dong, A., & Bhanu, B. (2005). Active concept learning in image databases. *IEEE Trans. Systems, Man and Cybernetics*, *35*(3), 450-466.

Erol, B., & Kossentini, F. (2005). Shape-based retrieval of video objects. *IEEE Trans. Multimedia*, *7*(1), 179-182.

Fletcher, L. A., & Kasturi, R. (1988). A robust algorithm for text string separation from mixed text/graphics images. *IEEE Trans. Pattern Analysis and Machine Intelligence*, *10*(6), 910-918.

Gonzalez, R. G., & Woods, R. E. (2002). *Digital Image Processing*. New Jersey, USA: Prentice Hall.

Hadjidemetriou, E., Grossberg, M. D., & Nayar, S. K. (2004). Multiresolution histograms and their use for recognition. *IEEE Trans. Pattern Analysis and Machine Intelligence*, *26*(7), 831-847.

Han, J., Ngan, K. N., & Zhang, H. J. (2005). A memory learning framework for effective image retrieval. *IEEE Trans. Image Processing*, *14*(4), 511-523.

Jing, F., Li, M., Zhang, H. J., & Zhang, B. (2004). An efficient and effective region-based image retrieval framework. *IEEE Trans. Image processing*, *13*(5), 699-709.

Joseph, S. H., & Pridmore, T. P. (1992). Knowledge-directed interpretation of mechanical engineering drawings. *IEEE Trans. Pattern Analysis and Machine Intelligence*, *14*(9), 928-940.

Lai, C. P., & Kasturi, R. (1994). Detection of dimension sets in engineering drawings. *IEEE Trans. Pattern Analysis and Machine Intelligence*, *16*(8), 848-855.

Li, J., & Wang, J. Z. (2004). Studying digital imagery of ancient paintings by mixtures of stochastic models. *IEEE Trans. Image Processing*, 12(3), 340-353.

Liapis, S., & Tziritas, G. (2004). Color and texture image retrieval using chromaticity histograms and wavelet frames. *IEEE Trans. Multimedia*, *6*(5), 676-686.

Lu, Y., & Lim, C. (2004). Information retrieval in document image databases. *IEEE Trans. Knowledge and Data Engineering*, *16*(11), 1398-1410.

Muneesawang, P., & Guan, L. (2004). An interactive approach for CBIR using a network of radial basis functions. *IEEE Trans. Multimedia*, *6*(5), 703-716.

Naqa, I. E., Galatsanos, N. P., & Wernick, M. N. (2004). A similarity learning approach to content-based image retrieval: application to digital mammography. *IEEE Trans. Medical Image*, *23*(10), 1233-1244.

Nezamabadi-pour, H., & Kabir, E. (2004). Image retrieval using histograms of un-color and bi-color blocks and directional changes in intensity gradient. *Pattern Recognition Letters*, *25*(14), 1547-1557.

Petrou, M., & Kadyrov, A. (2004). Affine invariant features from the trace transform. *IEEE Trans. Pattern Analysis and Machine Intelligence*, *26*(1), 30-44.

Prasad, B. G., Biswas, K. K., & Gupta, S. K. (2004). Region-based image retrieval using integrated color, shape, and location index. *Computer Vision and Image Understanding*, *94*(1-3), 193-233.

Swain, M. J., & Ballard, D. J. (1991). Color indexing. *Int. J. of Computer Vision*, *7*(1), 11-32.

Tsai, W. H. (1985). Moment-preserving thresholding: a new approach. *Computer Vision, Graphics, and Image Processing*, *29*(3), 377-393.

Vasconcelos, N., & Lippman, A. (2005). A multiresolution manifold distance for invariant image similarity. *IEEE Trans. Multimedia*, *7*(1), 127-142.

Vaxiviere, P., & Tombre, K. (1992). Celesstin: CAD conversion of mechanical drawings. *IEEE Trans. Computer*, *25*(7), 46-54.

Yin-Wen Huang received the B. S. and M. S. degrees in 2003 and 2005, both in computer science and engineering from Yuan Ze University, Taoyuan, Taiwan.

Mr. Huang's research interests include network management, color image processing and pattern recognition.

Chen-Chung Liu received the B. S. degree in electrical engineering, the M. S. and Ph. D. degrees in computer science and information engineering from Central University in Taiwan. He is currently an associate professor at Graduate Institute of Network Learning Technology, Central University. He is leading iLearn group (Interactive Technology and Learning) aiming to develop the interactive technologies including digital toys/games, personal handhelds and peripheral environment to support various learning scenarios, particularly those relating to the use of social intelligence that augments individual learning and creativity. Chen-Chung Liu is the recipient of 2006 Dr. Wu Da-You Research Award in Taiwan.

Shu-Yuan Chen received the B. S. degree in electrophysics in 1980, the M. S. and Ph. D. degrees both in computer engineering in 1982 and 1990, all from National Chiao Tung University, Hsinchu, Taiwan. Dr. Chen joined the faculty of Yuan Ze University (YZU), Taoyuan, Taiwan in 1994 and has been an YZU Professor in the Department of Computer Science and Engineering from 2002.

At YZU, Professor Chen has been the Head of the Department of Computer Science and Engineering from 2002 through 2004. Professor Chen's major research interests include image processing, pattern recognition, intelligent transportation systems, and image/video retrieval.