Multimed Tools Appl
DOI 10.1007/s11042-012-1233-7

Adaptive photograph retrieval method

**Hong-Bo Zhang · Shang-An Li · Shu-Yuan Chen ·
Song Zhi Su · Der-Jyh Duh · Shao Zi Li**

Abstract Access to electronic books, electronic journals, and web portals, which may contain graphics (drawings or diagrams) and images, is now ubiquitous. However, users may have photographs that contain graphics or images and want to access an electronic database to retrieve this information. Hence, an effective photograph retrieval method is needed. Although many content-based retrieval methods have been developed for images and graphics, few are designed to retrieve graphics and images simultaneously. Moreover, existing graphics retrieval methods use contour-based rather than pixel-based approaches. Contour-based methods, which are concerned with lines or curves, are inappropriate for images. To retrieve graphics and images simultaneously, this work applies an adaptive retrieval method. The proposed method uses histograms of oriented gradient (HOG) as pixel-based features. However, the characteristics of graphics and images differ, and this affects feature extraction and retrieval accuracy. Thus, an adaptive method is proposed that selects different HOG-based features for retrieving graphics and images. Experimental results demonstrate the proposed method has high retrieval accuracy even under noisy conditions.

Keywords Photograph retrieval · Graphics retrieval · Image retrieval · Histogram of oriented gradient · Pixel-based retrieval · Graphics/image classification

This work was partially supported by National Science Council of Taiwan, under Grant No. NSC-100-2221-E-155-086 and National Nature Science Foundation of China, under Grant No. 61202143.

H.-B. Zhang · S. Z. Su · S. Z. Li
School of Information Science and Technology, Xiamen University, Xiamen, Fujian, China

H.-B. Zhang · S. Z. Su · S. Z. Li
Fujian Key Laboratory of the Brain-like Intelligent Systems, Xiamen University, Xiamen, China

S.-A. Li · S.-Y. Chen (✉)
Department of Computer Science and Engineering, Yuan Ze University, Taoyuan, Taiwan
e-mail: cschen@saturn.yzu.edu.tw

D.-J. Duh
Department of Computer Science and Information Engineering, Chien Hsin University of Science and Technology, Taoyuan, Taiwan

Published online: 18 September 2012

1 Introduction

New computing technologies, media acquisition/storage devices, and multimedia compression standards have increased the amount of digital data generated and stored by computer users. Accessing electronic books, electronic journals, and web portals, which may contain graphics (drawings or diagrams) and images, is now easy. Particularly, users may have photographs that contain graphics or images and want to access an electronic database to retrieve this information.

In addition to the application of the e-book, e-journal and web retrieval, photography retrieval can be applied to content retrieval for e-learning system [7] and modality identification for medical images [12, 17]. In an e-learning system, class scenarios can be recorded in a video streaming format. Distance learning students can then review course content by watching a streaming e-learning video. Since important concepts and data are often presented using graphics and images, students could use the graphics and images from lecture videos as photographs to access information from an electronic database, ultimately increasing their learning efficiency. Thus, an effective photograph retrieval system is needed. Figure 1 shows examples of photographs from lecture videos [7].

Imaging modality is an important aspect of the image for medical retrieval [12, 17]. In user-studies, clinicians have indicated that modality is one of the most important filters to limit medical image search. Additionally, using the modality classification, the search results can be improved significantly. However, this modality is typically extracted from the caption and is often not correct or present. Studies have shown that the modality can be extracted from the image itself using visual features. The CLEF 2010 medical image retrieval task [12, 17] is to classify images into eight image modalities: computed tomography, magnetic resonance imaging, nuclear medicine, positron emission tomography, photography, radiography, ultrasound, and graphic (e.g. chart and drawing). The former seven categories are image modality and the last one is graphic modality. Therefore, the proposed adaptive photography retrieval can be used to facilitate CLEF 2010 medical image retrieval.

However, many content-based retrieval methods have been developed for images [9, 13, 21] and graphics [1–3, 5, 6, 8, 14–16, 18–20, 22], none are specifically designed for both graphics and images except our previous work [11]. The features used in our previous work include spatial histograms of binary pixel number, border length and gray level. The processed photographs is properly divided into two categories: graphics and image, type-

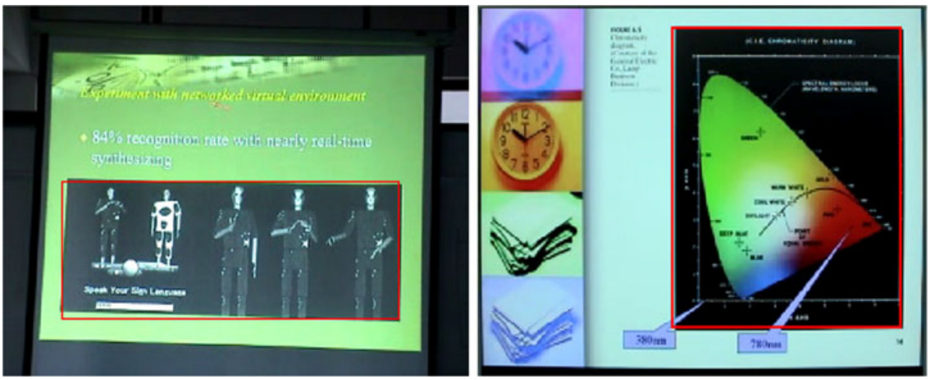


Fig. 1 Example photographs from lecture videos. The photographs are marked by *red rectangles*. (Courtesy of Chio *et al.* [7])

based matching is adopted to evaluate the similarity between the query photograph and each database prototype using different similarity measures according to the type of the query photograph which can be automatically determined according to the number of binary pixels in the entire photograph. However, our previous work is based on pixels themselves rather than oriented gradients of pixels in the proposed method. Since more information is involved, the proposed method is superior to our previous method as demonstrated by experiments.

Many content-based image retrieval (CBIR) methods [9, 13, 21] use color, texture, and shape features. On the other hand, existing graphics retrieval methods can be divided into two categories: pixel-based [3, 5, 8, 16, 19, 20, 22] and contour-based [1, 2, 6, 14, 15]. Pixel-based techniques consider all pixels or edge pixels within an image, while a contour-based method is concerned with lines or curves, which is inappropriate for images. The typical pixel-based techniques include Edge Pixel Density (EPD) [5], Shape Context (SC) [3], Scale-Invariant Feature Transform (SFIT) [16], Histograms of Oriented Gradients (HOG) [16], AHDH (Adaptive Hierarchical Density Histogram) [19, 22], and Structured Local Binary Haar Pattern (SLBHP) [20]. The typical contour-based techniques include Curvature Scale Space (CSS) [1], Inner Distance Shape Context (IDSC) [14], LC (Local Structure) [6, 15], and Triangle-Area-Representation (TAG) [2]. A hybrid of pixel-based and contour-based approaches was proposed in [18].

This work developed a novel and effective graphics/image retrieval approach that uses oriented gradient (HOG) as pixel-based features. However, the characteristics of graphics and images differ, which affects feature extraction and retrieval accuracy. In the proposed method, both the HOG and HOGE features for each database prototype are stored in repository, no matter the type of prototype is graphics or image. On the other hand, the query photograph type is first classified into two categories: graphics and image. The similarity between the query photograph and database prototype can then be assessed based on different features, *i.e.*, HOG for images, and HOGE for graphics. Thus, the proposed adaptive method has high retrieval accuracy. Experimental tests using collected databases demonstrate the superior accuracy of the proposed method.

The electronic database, 4000-database, has 2000 images and 2000 graphics (e.g., drawings or diagrams.) The graphics are collected from the patent database provided by Sidiropoulos *et al*. [19, 22], which includes binary patent images extracted from patent documents obtained from the European Patent Office [19, 22]. The images are mostly from the Corel database. The 200 images and 220 graphics were then chosen randomly from the electronic database and recorded with a digital camera to generate the query dataset of 420 photographs. Distortion and illumination vary in the query photographs.

To enlarge the image database with more diversity and general images and graphics, the 4000-database were extended to 5000-database. Moreover, the added 500 graphics were collected from an e-book with simple sketch rather than the original patent images with complex patterns; the added images were collected from Corel database but from classes which are hard to classify, for example different classes of landscape. Therefore, photography retrieval from the 5000-database is challenging. The other 50 graphics and 50 images were randomly chosen from the extended set and also recorded with a digital camera to generate the other query dataset of 100 photographs. Figures 2 and 3 present examples of database prototypes for 4000- and 5000-databases and query photographs for 420-query and 100-query datasets, respectively. Section 4 reveals the difficulty of 5000-database retrieval by experimental results.

The remainder of this paper is organized as follows. Section 2 presents the feature extraction and similarity measure methods based on local features. Section 3 addresses

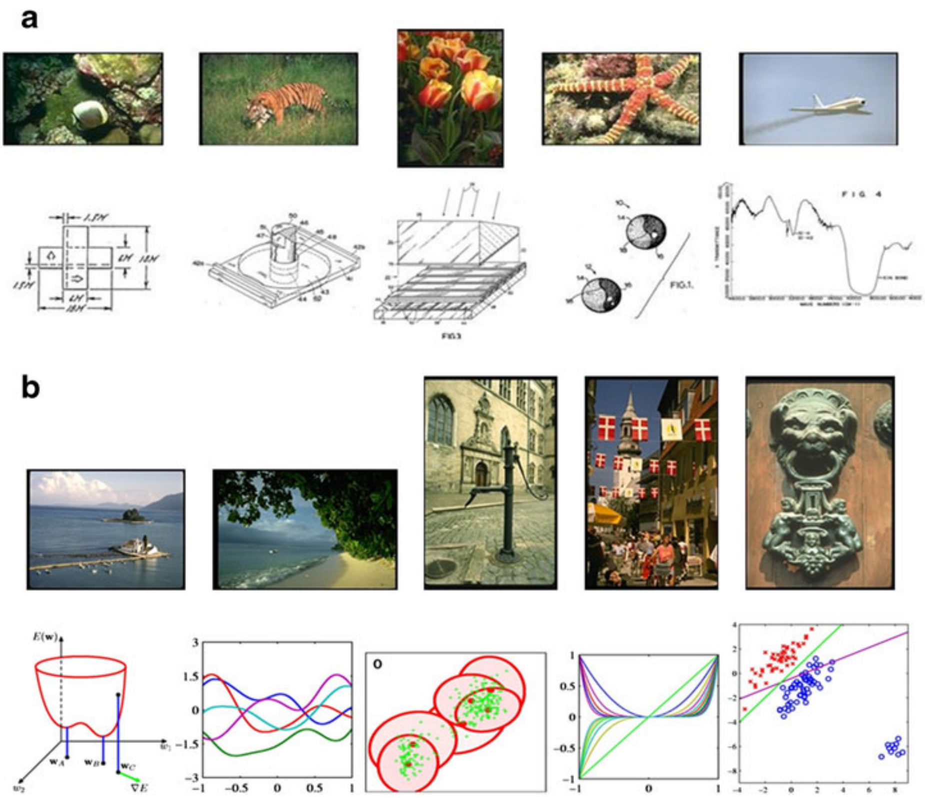


Fig. 2 Examples of database prototypes. **a** 4000-databse; (**b**) 5000-databse

graphics/image classification. Experimental results are given in Section 4. Finally, Section 5 gives conclusions and directions for future work.

2 Feature extraction

The proposed method uses histograms of gradients (HOGs) [8] as features. The query photographs are first preprocessed by a median filter to smooth noise in the photographs. In this work, default median filter size is 5×5. Both database prototypes and query photographs are then divided into small spatial regions/blocks from which the HOG is computed. The HOGs, one for each block, are then concatenated to form the representation for graphics and images. The similarity between a query photograph and database prototype is then computed as the chi-square distance measure based on HOGs. The retrieved list has similar graphics/images ranked by distance values. Figure 4 shows the framework of the proposed method.

Traditional block-partition divides an entire image into a grid, in which all blocks are the same size. Notably, block size is crucial because a large block may enclose a contiguous region and produce conspicuous features, whereas a small block cannot adequately represent object characteristics. Rectangle partitioning is the most common method for representing small spatial regions in an image. An image can be divided into several rectangles of the

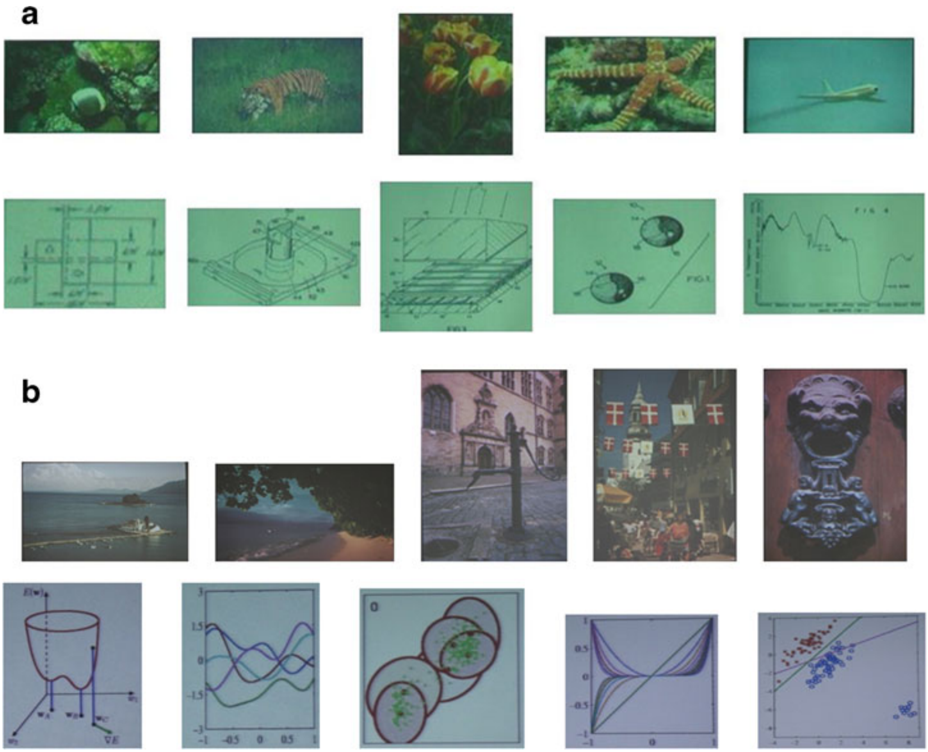


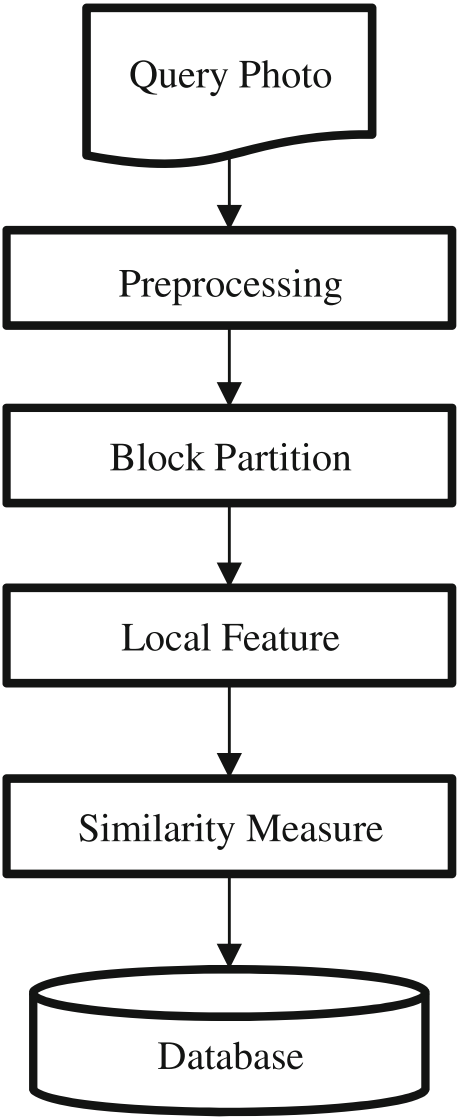
Fig. 3 Examples of query photograph. **a** 420-query dataset; (**b**) 100-query dataset

same size. The ratio of block size to image size generally depends on the total number of blocks. In other words, if an image is divided into $M \times N$ blocks, block size is $h/M \times w/N$, where h and w are image height and width, respectively. In this work, both M and N are set to 9.

The HOG, first developed for use in human detection [8], divides image windows into small spatial regions called cells. A local histogram of gradient direction over pixels in the cell is then constructed. The most common gradient computation method is to apply a mask in both the horizontal and vertical directions. This work uses two masks ([−1,0,1] and [1,0,−1]) to filter intensity data of an image to obtain the orientation (or angle) of the current pixel. Each pixel within a cell then casts a weighted vote for an orientation-based histogram channel based on values calculated by gradient computation (Fig. 5). Notably, histogram channels are evenly spread over 0–180° or 0–360°. In this work, angles of 0–180° are divided into ten 18° intervals. To increase tolerance for vertical and horizontal angles, angles of 0–9° degrees and 171–180° are set to the same interval; angles of 81–99° form a new interval (Fig. 6). The Histogram of Oriented Gradients of Edge Pixels (HOGEs) is a modified version of HOG achieved by extracting the HOGs from an edge image. That is, before extracting the HOGs, the Canny edge detector [4] is applied. Notably, HOGE is insensitive to illumination. Each HOGE histogram also has a dimension of 10.

After partitioning, feature extraction is applied to construct a local feature histogram for each block, which is then concatenated to form the graphics/image representation (Fig. 7). The concatenated histogram is a graphics/image representation. To rank retrieval results, the χ^2 distance is calculated to determine the similarity between two graphics/images using the concatenated histograms. For a consistent similarity measure, each value for bin i, $h(i)$, is

Fig. 4 Framework of the proposed method



normalized to $h'(i)$ within the range of 0–1 by the following equation:

$$h'(i) = \frac{h(i)}{\sum_{i=1}^{n} h(i)} \tag{1}$$

where n is the total number of bins, *i.e.*, 10 in this work. The χ^2 distance is then calculated using Eq. (2), where $h_1^{'}(i)$ and $h_2^{'}(i)$ are two different histograms, and i is the corresponding bin number.

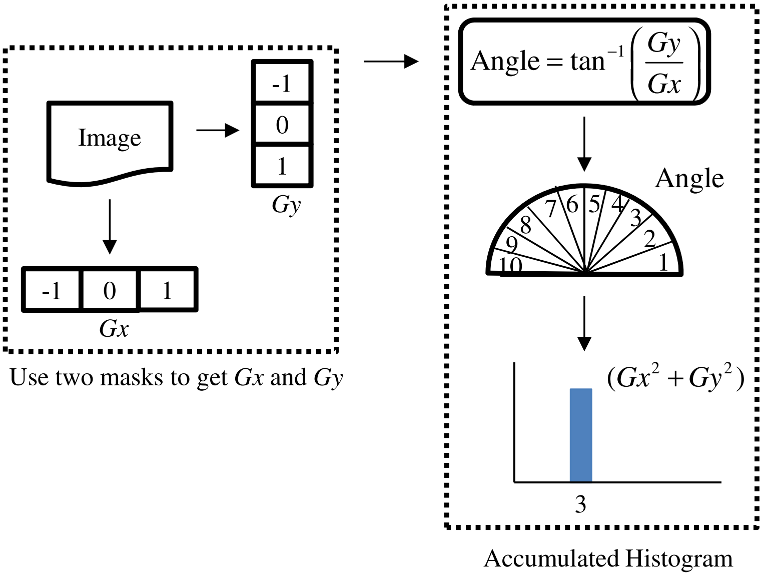


Fig. 5 HOG feature extraction

$$\chi^2\, distance = \sum_{i=1}^{n} \left(\frac{h_1'(i) - h_2'(i)}{h_1'(i) + h_2'(i)} \right)^2 \tag{2}$$

Notably, as the χ^2 distance decreases, similarity of the two graphics/images increases.

3 Graphics/image classification

The characteristics of graphics and images differ, and this affects feature extraction and, in turn, retrieval accuracy. Table 1 show retrieval accuracies using HOG and HOGE individually; HOG and HOGE are appropriate for dealing with images and graphics, respectively. Therefore, to improve the accuracy of graphics/image retrieval, the preliminary step is to classify the query photograph type. The similarity between a query photograph and database prototype can then be assessed based on different features, *i.e.*, HOG for images, and HOGE for graphics. Note that type classification is performed on only query photograph. On the other hand, both the HOG and HOGE features for each database prototype are stored in repository, no matter the type of prototype is graphics or image. Figure 8 shows the framework of the proposed photo retrieval approach.

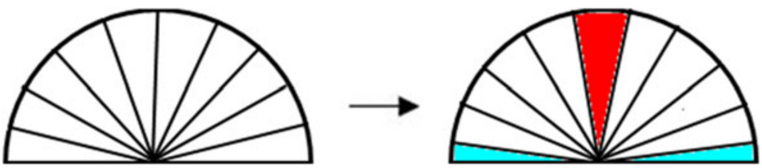


Fig. 6 Modification of angle interval

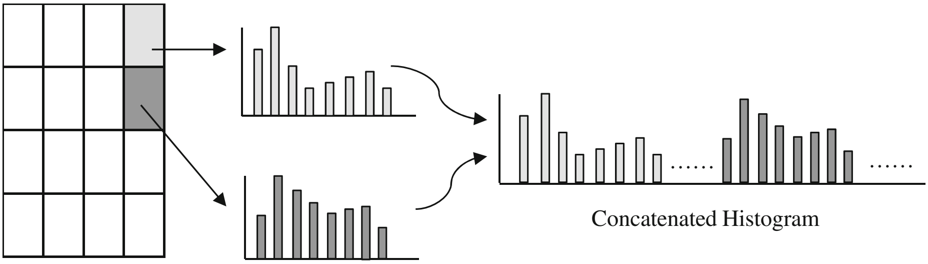


Fig. 7 Concatenated histogram

The proposed graphics/image classification algorithm uses standard deviations and entropy of a grayscale distribution to classify photographs into two categories. Generally, graphics contain many abrupt pattern changes, such as lines, curves, and nodes, over a uniform background; while images often have many details, smooth color changes from one part of an image to the other, and a variety of rich textures. Hence, graphics should have low standard deviations in the foreground and background, while images should have higher standard deviations. Moreover, graphics should have low entropy in the entire photograph, while images should have higher entropy.

The K-means algorithm [10] with K=2 is applied to divide the query photograph into two sets, foreground and background, according to the grayscale distribution. Note that the function of 2-means algorithm is just a binary thresholding to binarize the photograph. In this work, the two initial means are set to maximum and minimum values in the photographs grayscale histogram. Let sets F and B be the foreground and background sets obtained by the two-means algorithm for each photograph, respectively. Let each pixel (x,y) in the photograph have the gray value $g(x,y)$. Equations (3) and (4) are applied to calculate standard deviations s_f and s_b for foreground F and background B, respectively.

$$s_f = \sqrt{\frac{\sum_{\forall(x,y)\in F}\left[\mu_f - g(x,y)\right]^2}{N_f}}, \mu_f = \frac{\sum_{\forall(x,y)\in F} g(x,y)}{n_f}, n_f = \sum_{\forall(x,y)\in F}(1) \tag{3}$$

$$s_b = \sqrt{\frac{\sum_{\forall(x,y)\in B}\left[\mu_b - g(x,y)\right]^2}{n_b}}, \mu_b = \frac{\sum_{\forall(x,y)\in B} g(x,y)}{n_b}, n_b = \sum_{\forall(x,y)\in B}(1) \tag{4}$$

where n_f, μ_f, and s_f are the number of pixels, the mean of grayscale, and the standard deviation of grayscale in the foreground set. The same definition is for n_b, μ_b, and s_b for the background set. Let decision standard deviation s_d be the sum of s_f and s_b. Figure 9 shows

Table 1 Retrieval accuracies using various features (Median filter size=5×5)

Method	HOG	HOGE
Image	97.00 %	75.00 %
Graphics	89.55 %	92.27 %
Total	93.00 %	84.04 %

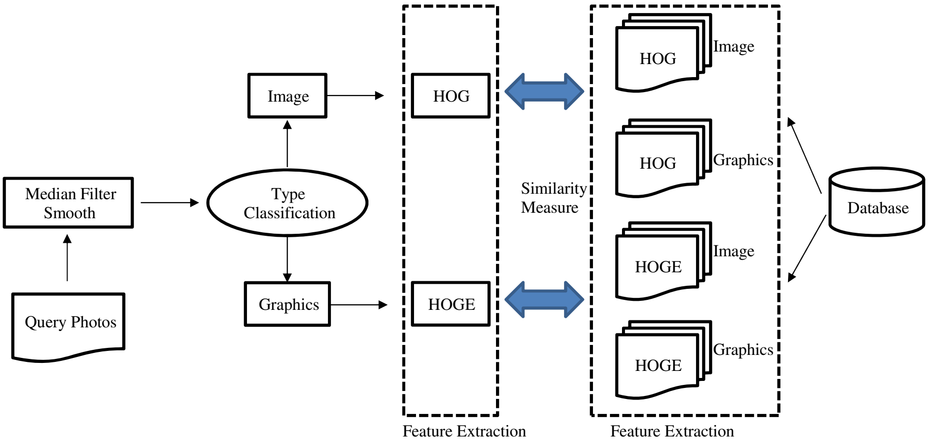


Fig. 8 Framework of photograph retrieval method

the distributions of s_f, s_b, s_d for the 200 images and 220 graphics. It is obvious that graphics have low s_d, while images have high s_d.

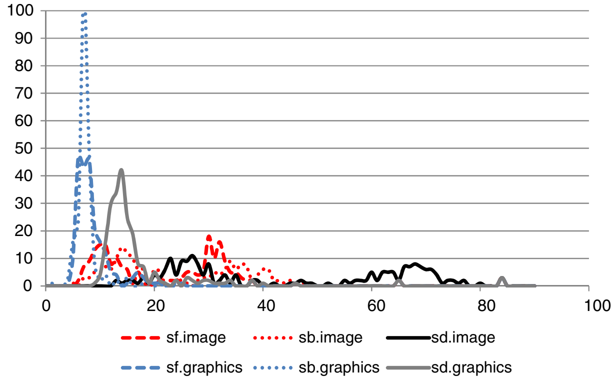
On the other hand, the entropy E of each photograph is calculated as follows. The grayscales of the photograph are first quantized into 10 levels. The normalized grayscale histogram in terms of 10 bins, $h_g(i), i = 1, \cdots, 10$, of the entire photograph, can then be obtained. Finally, the entropy E of each photograph is computed by the following equation

$$E = -\sum_{i=1}^{m} h_g(i) \ln\left(h_g(i)\right) \tag{5}$$

where m is the number of bins, *i.e.*, 10 in this work. Figure 10 shows examples of normalized grayscale histograms of image-type and graphics-type photographs, respectively. Figure 11 shows the distributions of the entropy E for the 200 images and 220 graphics. It is obvious that graphics have low entropy, while images have higher entropy.

The proposed graphics/image classification algorithm uses standard deviations and entropy of grayscale distribution to classify photographs into two categories, graphics and images. Different HOG-based features for retrieving graphics and images are then selected

Fig. 9 Distributions of s_f, s_b, and s_d



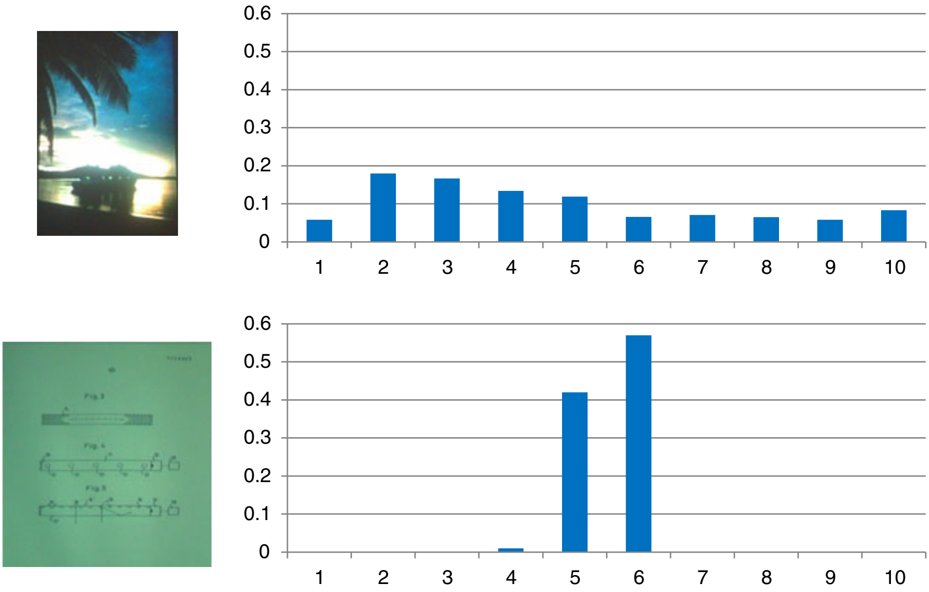


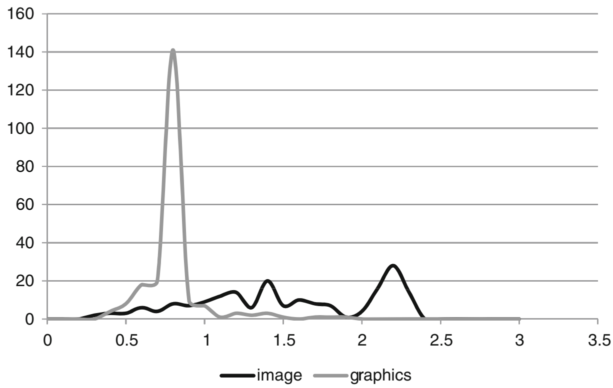
Fig. 10 Examples of normalized grayscale histograms of photographs. The image and graphics are shown in the first and the second rows, respectively

accordingly. If s_d is smaller than the threshold T_s or E is smaller than the threshold T_e, the query photograph is graphics, otherwise the query photograph is image. Threshold values of T_s and T_e are empirically set to 15 and 1 in this work. Figure 12 shows the framework of the proposed classification method.

4 Experimental results

The proposed method was implemented on a GIGABYTE motherboard with a quad 2.66 GHz core Intel Q8400 CPU and 4 Gigabytes of DDR2 SDRAM. The operating system was Microsoft Windows XP Service Pack 3. The program was developed in C++ language

Fig. 11 Distributions of entropy



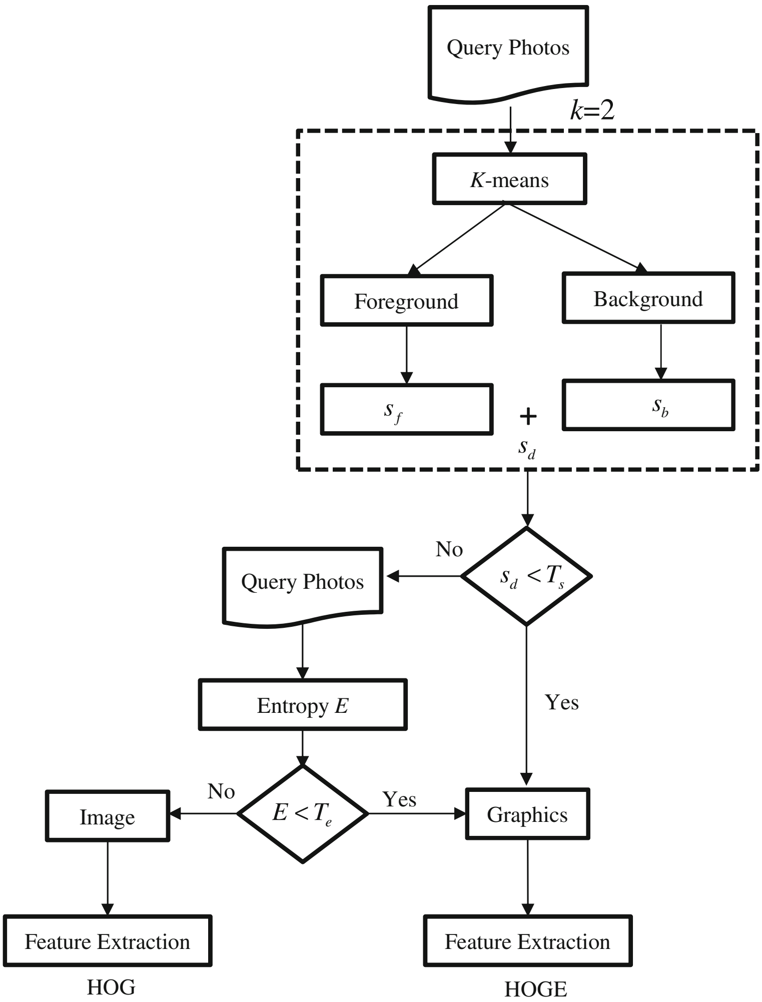


Fig. 12 Framework of photograph classification

with an open source Open CV library and compiled under Microsoft Visual Studio 2010. The block partition is a rectangle partition with 9×9 blocks. The median filter is 5×5.

The experimental procedure is described as follows. For each query photograph in the query set, media filter is first performed. The standard deviation and entropy of the photograph are then computed to classify its type. If the type of the photograph is graphics, the χ^2 distance between the query photograph and all the database prototypes, no matter graphics or image, in terms of HOGE features are computed and the prototypes are ranked according to the respective χ^2 distance in the increasing order. If the corresponding prototype is ranked first, the retrieval result for the photograph is considered correct. Similarly, if the type of photograph is classified as image, the procedure is the same except adopting HOG features instead of HOGE. Finally, the precision accuracy is defined as the ratio of the number of retrieved corresponding prototypes over the number of query photographs, *i.e.*, the number of photographs in the query set.

4.1 Effectiveness evaluation

Photograph retrieval performance is measured as retrieval accuracy. Retrieval accuracy is computed as the ratio of the number of graphics/images correctly retrieved to the total number of queries. Table 2 shows retrieval accuracies using different strategies. The retrieval method using adaptive features (*i.e.*, HOGE or HOG) is better than that using only HOGE or HOG. The conditional accuracies for image-type and graphic-type photographs using the adaptive method, HOG, and HOGE are 95.50 %, 97.00 %, and 75.00 % and 92.27 %, 89.55 %, and 92.27 %, respectively (Table 2). Hence, differences in accuracy between image-type photographs and graphics-type photographs using the adaptive method, HOG, and HOGE are 3.23 %, 7.45 %, and 7.27 %, respectively. This observation implies that the difference in retrieval accuracies between images and graphics using the adaptive strategy is smaller than that using HOG or HOGE alone. Restated, the adaptive method is good for both images and graphics; while HOG and HOGE are best for images and graphics, respectively.

The SIFT-matching [16] was implemented in this study for comparison. The reasons for adopting SIFT-matching are as follows. First, to the best of our knowledge, there are no other methods except our previous work [11] to retrieve graphics and images simultaneously. Second, contour-based methods, which are concerned with lines and curves, are inappropriate for images. Third, SIFT-matching is well-known and commonly used. Assume that N_q and N_p SIFT interest points are detected in the query photo Q and the database prototype P, respectively. The similarity for each pair of N_q and N_p SIFT interest points is calculated using Euclidean distance in terms of the 128 dimensional SIFT feature vector. The smaller is the distance, the higher is the similarity. The one with the smallest Euclidean distance D is regarded as the matched point for the interest point in the photograph provided that D is less than a threshold T_m. Let the number of matched points be N_m. The similarity between Q and P is then defined by the following equation

$$Similarity(Q, P) = \frac{N_m}{N_q} + \frac{N_m}{N_p} \tag{6}$$

The retrieval results for the query photo Q can then be ranked in the decreasing order of *Similarity* (Q, P). In this study, T_m is set to 0.3 empirically. The accuracy of using SIFT-matching for the 420 queries on 4000-database is listed in Table 3.

In addition, our previous work [11] was also applied on the 4000-database. The accuracy of our previous work is listed in Table 3. Our previous work was implemented by adopting 9×9 partition as in this work. The features used in our previous work are also histograms but in terms of binary pixel numbers, border length and gray levels. The type classification is based on the

Table 2 Retrieval accuracies using various strategies (Median filter size=5×5) for 420 queries. The highest accuracies in each row are marked as bold

	Classification	Adaptive	HOG	HOGE
Image	Image(195)	**96.92 %**	**96.92 %**	75.90 %
	Graphics(5)	40.00 %	**100.00 %**	40.00 %
	200	95.50 %	**97.00 %**	75.00 %
Graphics	Image(68)	**95.59 %**	**95.59 %**	**95.59 %**
	Graphics(152)	**90.79 %**	86.84 %	**90.79 %**
	220	**92.27 %**	89.55 %	**92.27**%
Total	420	**93.81 %**	93.00 %	84.04 %

Table 3 Comparison of retrieval accuracies using various methods (Median filter size=5×5) for 420 queries. The highest accuracies in each row are marked as bold

Method	Adaptive	HOG	HOGE	SIFT	[20]
Image	95.50 %	**97.00 %**	75.00 %	86.50 %	59.00 %
Graphics	**92.27 %**	89.55 %	**92.27 %**	35.00 %	77.73 %
Total	**93.81 %**	93.00 %	84.04 %	59.52 %	68.81 %

entire ratio of binary pixels, denoted as R. In general, the ration of binary pixels of graphics is small. Therefore, the type of photograph is classified as graphics, hybrid, and image if R is less than 0.25, between 0.25 and 0.5, and over 0.5, respectively. The similarity measure is defined in terms of histograms of binary pixel number and border length, histograms of binary pixel number, border length and gray levels, and histograms of gray levels, for the type of graphics, hybrid, and image, respectively. The details can be referred to [11]. Table 3 shows that the proposed adaptive method is superior to SFIT-matching and our previous work.

The proposed method was also applied to 5000-database using the other 100-query dataset. The same conclusions can be drawn (Tables 4 and 5). The 100-query dataset on 5000-database is more challenging than 420-query dataset on 4000-databse can be verified by experimental results. Tables 3 and 5 reveal that the accuracies of our method are 93.81 % and 82.00 % for 420-query dataset and 100-query dataset, respectively. In addition, the accuracies of SFIT-matching are 59.92 % and 51.00 % for 420-query dataset and 100-query dataset, respectively; the accuracies of our previous work [11] are 68.81 % and 51.00 % for 420-query dataset and 100-query dataset, respectively. Therefore, the accuracies degrade from 420-query dataset on 4000-databse to 100-query dataset for 5000-database no matter the used methods. Nevertheless, the accuracies of the proposed method are higher than those of using SIFT-matching and our previous method. Therefore, the proposed adaptive method works effectively for the diversity of databases.

However, the adaptive strategy works more effectively on 5000-database than on 4000-database. Table 3 shows that the proposed adaptive method on 4000-database has accuracy 93.81 % which is higher than 93.00 % using only HOG feature and 84.04 % using only HOGE; while Table 5 shows that the proposed adaptive method on 5000-database has accuracy 82.00 % which is higher than 70.00 % using only HOG feature and 69.00 % using only HOGE. The adaptive strategy improves the accuracy rates 5.29 % and 12.5 % for the 4000-database and the 5000-database, respectively. Figure 13 shows retrieval results using adaptive features for the 420-query and 100-query datasets, respectively.

Table 4 Retrieval accuracies using various strategies (Median filter size=5×5) for 100 queries. The highest accuracies in each row are marked as bold

	Classification	Adaptive	HOG	HOGE
Image	Image(42)	**100.00 %**	**100.00 %**	50.00 %
	Graphics(8)	37.50 %	**87.50 %**	37.50 %
	50	90.00 %	**98.00 %**	48.00 %
Graphics	Image(16)	31.25 %	31.25 %	**81.25 %**
	Graphics(34)	**94.12 %**	47.06 %	**94.12 %**
	50	74.00 %	42.00 %	**90.00 %**
Total	100	**82.00 %**	70.00 %	69.00 %

Table 5 Comparison of retrieval accuracies using various methods (Median filter size=5×5) for 100 queries. The highest accuracies in each row are marked as bold

Method	Adaptive	HOG	HOGE	SIFT	[20]
Image	90.00 %	**98.00 %**	48.00 %	80.00 %	14.00 %
Graphics	74.00 %	42.00 %	**90.00 %**	22.00 %	88.00 %
Total	**82.00 %**	70.00 %	69.00 %	51.00 %	51.00 %

4.2 Robustness evaluation

To prove that the proposed adaptive method is insensitive to parameter settings, retrieval accuracies under different parameter values are tested. The parameters tested were type classification threshold, T_s, Canny edge operator threshold value, C, and median filter size. In this experiment, the feasible parameter values are T_s=15 and 20, and C=80 and 65, and median filter size is 5×5 and 3×3 and no median filters. Notably, as the Canny edge threshold value C decreases, more edge details are obtained. Figure 14 shows retrieval accuracies and accuracy differences between images and graphics under different parameter values. For the median filter sized 5×5 and 3×3 and no median filters, the best parameter settings are T_s=15, 15, and 20 and C=65, 80, and 65, generating the highest accuracies (Fig. 14). However, regardless of parameter values, the accuracies with the adaptive strategy exceed 89.00 %, and the difference in retrieval accuracies between images and graphics using the adaptive strategy is smaller than that when using the HOG or HOGE.

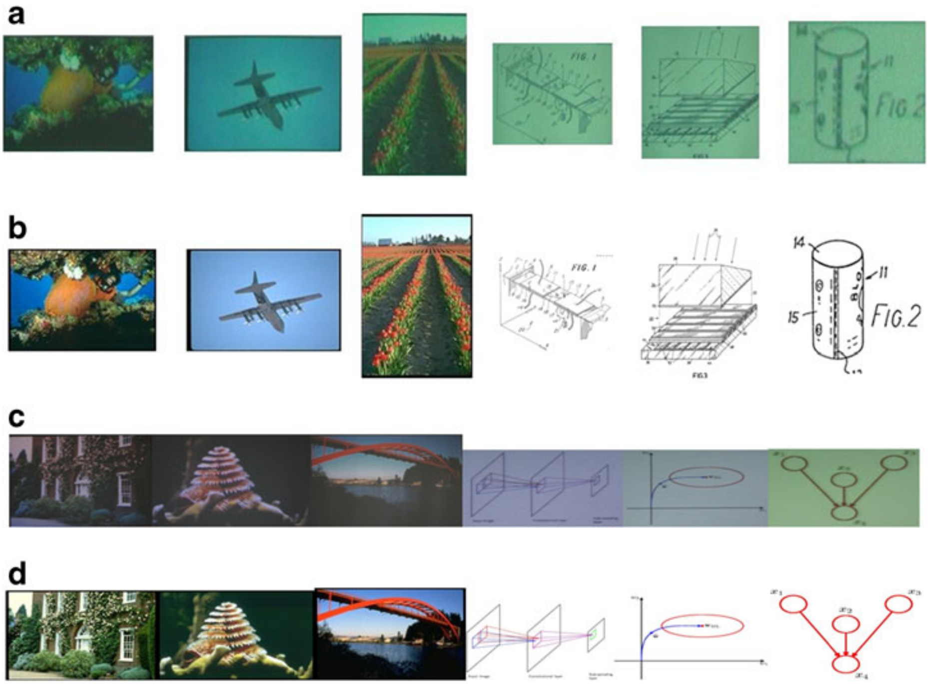


Fig. 13 Example retrieval results. **a** and **b** Query photographs and retrieval results for 420-query dataset; (**c**) and (**d**) Query photographs and retrieval results for 100-query dataset

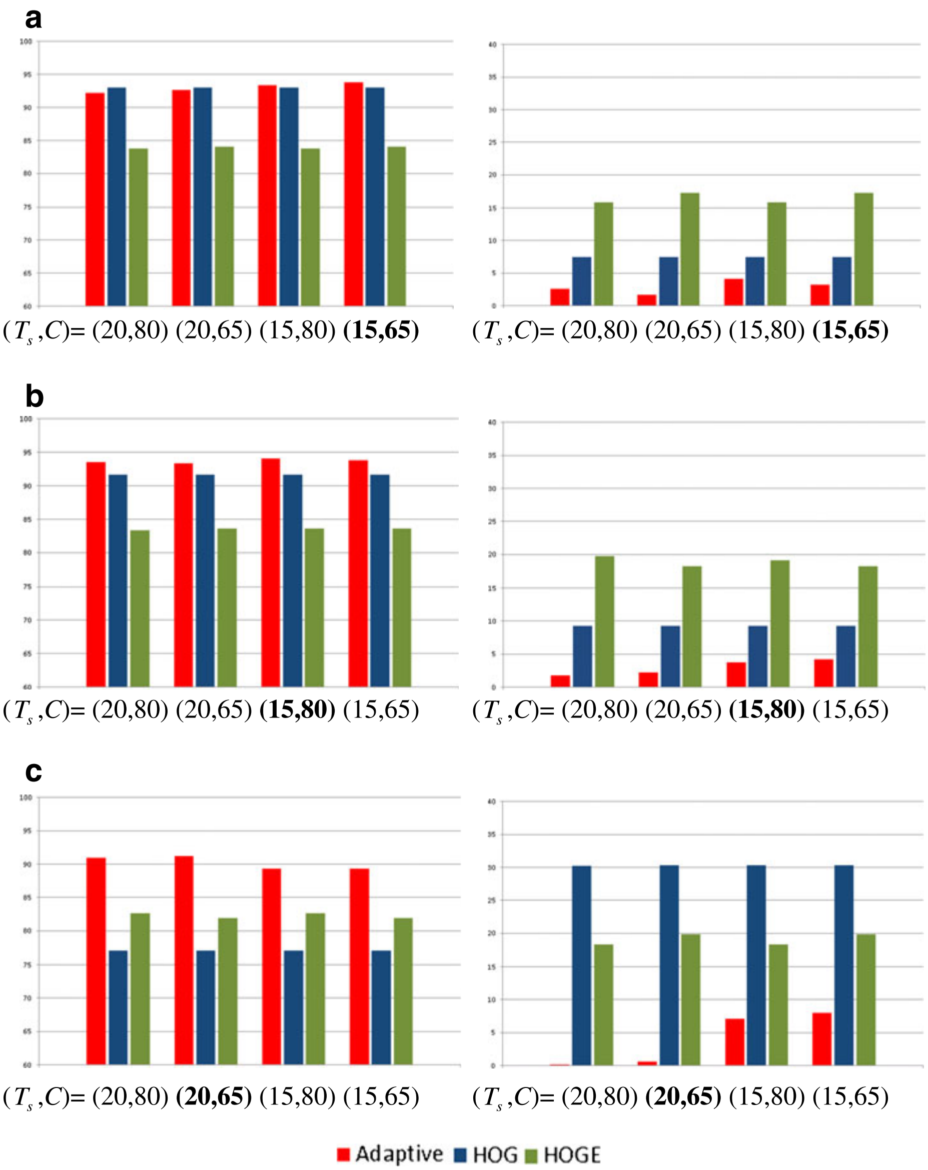


Fig. 14 Retrieval accuracies under various parameter values. Left and right columns represent retrieval accuracy and differences in accuracy between images and graphics, respectively. **a**, **b**, and **c** are median filter sizes of 5×5 and 3×3 and no median filter

Noise was added to query photographs to demonstrate the robustness of the proposed method. Noise, such as Gaussian noise, salt-and-pepper noise, and Gaussian blurring, were added to photographs. Figure 15 shows some noisy photographs. Table 6 lists retrieval accuracies under various types of noise. Notably, for the median filter sized 5×5and 3×3 and no median filters, parameter values were set at T_s=15, 15 and 20 and C=65, 80, and 65, respectively, since these values generate the highest accuracies. The median filter sized 5×5

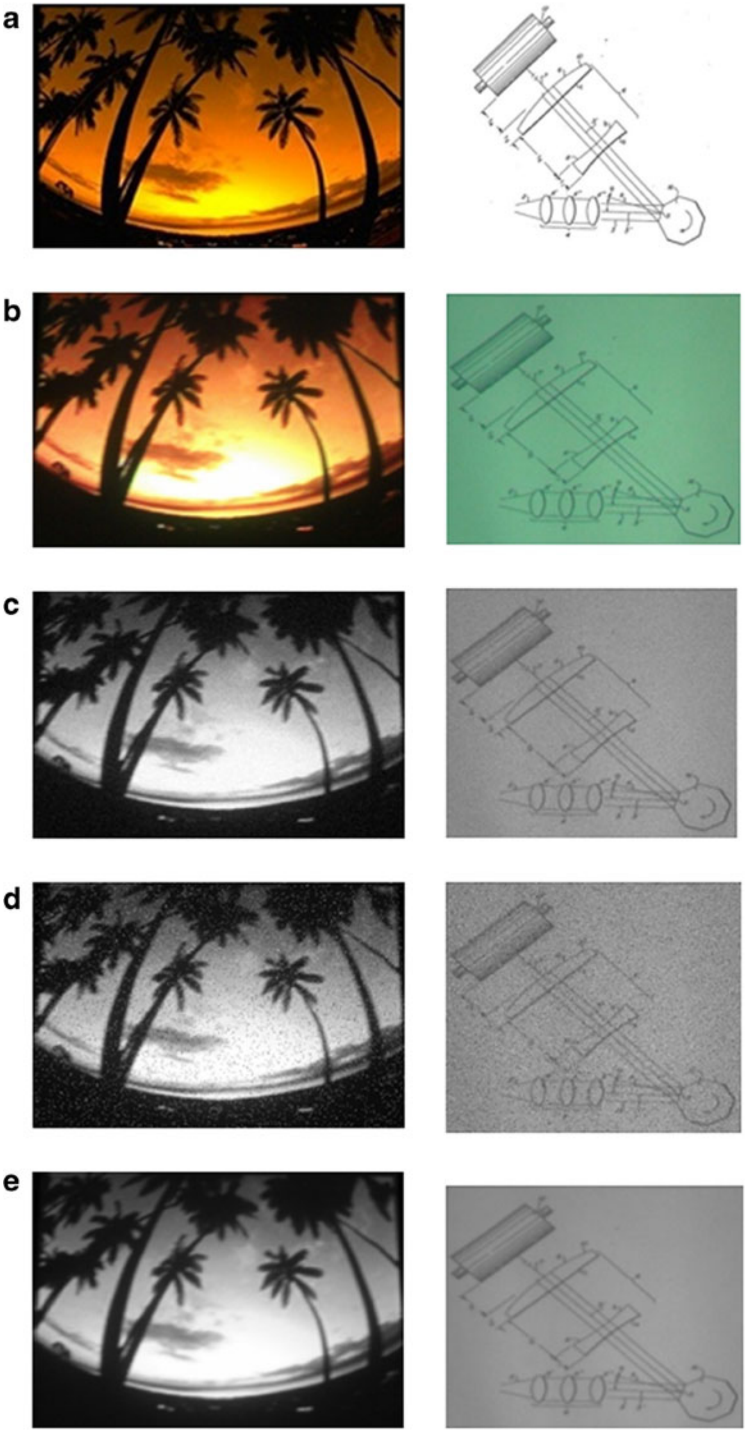


Fig. 15 Examples of noise. **a** Electronic prototypes; (**b**) original photographs; (**c**) Gaussian noise; (**d**) salt-and-pepper noise; and (**e**) Gaussian blurring

Table 6 Retrieval accuracies under noise. The highest accuracies in each row are marked as bold

Noise Type	Photo Type	Median filter of 3×3 (T,C)=(15,80)	Median filter of 5×5 (T,C)=(15,65)	No median filters (T,C)=(20,65)
Original	Image(200)	**96.00 %**	95.50 %	91.50 %
	Graphics(220)	**92.27 %**	**92.27 %**	90.91 %
	Total(420)	**94.05 %**	93.81 %	91.19 %
Gaussian Noise (Sigma5)	Image(200)	93.50 %	**96.00 %**	60.00 %
	Graphics(220)	**90.91 %**	**90.91 %**	80.91 %
	Total(420)	92.14 %	**93.33 %**	70.95 %
Gaussian Noise (Sigma10)	Image(200)	70.50 %	**91.50 %**	24.50 %
	Graphics(220)	57.73 %	**84.09 %**	5.91 %
	Total(420)	63.81 %	**87.62 %**	14.76 %
Gaussian Noise (Sigma15)	Image(200)	48.00 %	**78.50 %**	7.00 %
	Graphics(220)	10.45 %	**68.18 %**	3.64 %
	Total(420)	28.33 %	**73.10 %**	5.24 %
Gaussian Noise (Sigma20)	Image(200)	34.50 %	**59.00 %**	1.00 %
	Graphics(220)	**5.91 %**	4.09 %	1.36 %
	Total(420)	19.52 %	**51.19 %**	1.19 %
Salt-and-Pepper (0.02)	Image(200)	**96.00 %**	**96.00 %**	1.00 %
	Graphics(220)	**92.27 %**	91.82 %	1.36 %
	Total(420)	**94.05 %**	93.81 %	1.19 %
Salt-and-Pepper (0.05)	Image(200)	**96.50 %**	96.00 %	0.00 %
	Graphics(220)	**92.27 %**	91.36 %	0.00 %
	Total(420)	**94.29 %**	93.57 %	0.00 %
Gaussian Blurring (3×3)	Image(200)	**96.00 %**	**96.00 %**	94.50 %
	Graphics(220)	**93.64 %**	**93.64 %**	92.73 %
	Total(420)	**94.76 %**	**94.76 %**	93.57 %
Gaussian Blurring (5×5)	Image(200)	**96.00 %**	**96.00 %**	94.00 %
	Graphics(220)	90.91 %	92.27 %	**93.18 %**
	Total(420)	93.33 %	**94.05 %**	93.57 %
Gaussian Blurring (7×7)	Image(200)	95.00 %	**95.50 %**	94.00 %
	Graphics(220)	82.27 %	87.27 %	**90.91 %**
	Total(420)	88.33 %	91.19 %	**92.38 %**

is the most robust for noise (Table 6). Thus, the median filter sized 5×5 is the default filter in this work.

5 Conclusions

An effective photograph retrieval method using pixel-based features is used to retrieve graphics and images adaptively. The proposed method adopts HOG-based features to represent photographs. However, the characteristics of graphics and images differ, and this affects feature extraction results, and subsequently retrieval accuracy. Thus, a type classification is proposed to classify the type of photograph according to its standard deviation and entropy. Henceforth, the proposed adaptive method selects HOG or HOGE features for the photograph to retrieve corresponding prototype according to the type being image or graphics. Thereby, the proposed adaptive method is superior to SFIT-matching and our previous work, as demonstrated by experimental results. Future research can be directed to extend the proposed graphics/image retrieval method for partial matching and integrate the proposed graphics/image retrieval method with an illustration extraction method to construct an illustration retrieval system for lecture videos.

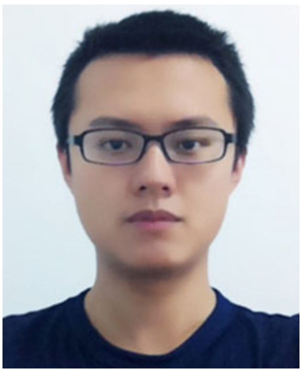
Acknowledgement The authors would like to thank P. Sidiropoulos, S. Vrochidis, and I. Kompatsiarisa for providing patent image database and the anonymous reviewers for the valuable and insightful comments on the earlier version of this manuscript.

References

1. Abbasi S, Mokhtarian F, Kittler J (1999) Curvature scale space image in shape similarity retrieval. Multimed Syst 7(6):467–476
2. Alajlan N, Kamel M, Freeman G (2008) Geometry-based image retrieval in binary image databases. IEEE Trans Pattern Anal Mach Intell 30(6):1003–1013
3. Belongie S, Malik J, Puzicha J (2002) Shape matching and object recognition using shape contexts. IEEE Trans Pattern Anal Mach Intell 24(4):705–522
4. Canny JF (1986) A computational approach to edge detection. IEEE Trans Pattern Anal Mach Intell 8 (6):679–714
5. Chalechale A, Naghdy G, Mertins A (2005) Sketch-based image matching using angular partitioning. IEEE Trans Syst Man Cybern Syst Hum 35(1):28–41
6. Chi Y, Leung MKH (2007) ALSBIR: a local-structure-based image retrieval. Pattern Recogn 40(1):244–261
7. Chio J-W, Chen S-Y (2012) Illustration extraction from video streams. J Pattern Recogn Res 7(1):56–71
8. Dalal N, Triggs B (2005) Histograms of oriented gradients for human detection. Proc IEEE Conf Comput Vision Pattern Recogn 1:886–893
9. Datta R, Joshi D, Lia J, Wang JZ (2008) Image retrieval: ideas, influences, and trends of the new age. ACM Comput Surv 40(2):1–60
10. Duda RO, Hart PE, Stork DG (2001) *Pattern Classification*, 2nd, Wiley
11. Huang Y-W, Liu C-C, Chen S-Y (2007) Graph/image legend retrieval. Asian J Health Inform Sci 2, nos. 1–4:79–102
12. Kahn CE Jr, Kalpathy-Cramer J, Lam CA, Eldredge CE (2012) Accurate determination of imaging modality using an ensemble of text- and image-based classifiers. J Digital Imag 25(1):37–42
13. Lew MS, Sebe N, Djeraba C, Jain R (2006) Content-based multimedia information retrieval: state of the art and challenges. ACM Trans Multimed Comput Commun Appl 2(1):1–19
14. Ling H, Jacobs D (2007) Shape classification using the inner-distance. IEEE Trans Pattern Anal Mach Intell 29(2):286–299
15. Liu R, Wang Y, Baba T, Masumoto D (2010) Shape detection from line drawings with local neighborhood structure. Pattern Recogn 43(5):1907–1916
16. Lowe DG (2004) Distinctive image features from scale-invariant keypoints. Int J Comput Vis 60(2):91–110
17. Müller H, Kalpathy-Cramer J, Eggel I, Bedrick S, Kahn CE Jr, Hersh W, "Overview of the CLEF 2010 medical image retrieval track," http://clef2010.org/resources/proceedings/ImageCLEF2010_medOverview
18. Qi H, Li K, Shen Y, Qu W (2010) An effective solution for trademark image retrieval by combining shape description and feature matching. Pattern Recogn 43(6):2017–2027
19. Sidiropoulos P, Vrochidis S, Kompatsiarisa I (2011) Content-based binary image retrieval using the adaptive hierarchical density histogram. Pattern Recogn 44(4):739–750

20. Su S-Z, Chen S-Y, Li S-Z, Li S-A, Duh D-J (2010) Structured local binary Haar pattern for pixel-based graphics retrieval. Electron Lett 46(14):996–998
21. Torralba A, Fergus R, Freeman WT (2008) Eighty million tiny images: a large dataset for non-parametric object and scene recognition. IEEE Trans Pattern Anal Mach Intell 30(11):1958–1970
22. Vrochidis S, Papadopoulos S, Moumtzidou A, Sidiropoulos P, Pianta E, Kompatsiaris I (2010) Towards content-based patent image retrieval: a framework perspective. World Patent Inform 32(2):94–106

Hong-Bo Zhang received the B.S. degree in Computer Science and Technology from Shenyang Normal University, Shenyang, China, in 2008. He is currently working toward Ph.D. degree in artificial intelligent in School of Information Science and Technology of Xiamen University, China. During Sept. 2011 and Jan. 2012, he went to Yuan Ze University as an exchange student. His research interests include human action analysis, object detection/recognition, and image/video retrieval.



Shang-An Li received the B.S. degree in Computer and Communication Engineering from Ming Chuan Normal University, Taoyuan, Taiwan, in 2008 and the M.S. degree in Computer Science and Engineering from Yuan Ze University, Taoyuan, Taiwan, in 2010. Since 2010, he has served in military. His research interests include network management, image processing and pattern recognition.

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 to 2004. Professor Chen's major research interests include image processing, pattern recognition, intelligent transportation systems, and image/video retrieval.

Song-Zhi Su received the B.S. degree in Computer Science and Technology from Shandong University, China, in 2005. He received M.S. and Ph.D degree in Computer Science in 2008 and 2011, both from Xiamen University, Fujian, China. He joined the faculty of Xiamen University as an assistant professor in 2011. His research interests include pedestrian detection, time-of-flight camera based human action recognition and image/video retrieval.

Der-Jyh Duh received the B.S. degree in Electrophysics from National Chiao Tung University in 1981, the M.S. degrees in Electrical Engineering from Tatung University in 1983 and the Ph.D. degree in Computer Science and Engineering from Yuan Ze University in 2005. He has been working as a senior system engineering at the Information and Communication Research Division of Chung Shan Institute of Science and Technology (CSIST) from 1983 to 2008. Dr. Duh joined the faculty of Chien Hsin University of Science and Technology, (CHU), Taoyuan, Taiwan in 2008 and has been an assistant professor in the Department of Computer Science and Information Engineering from 2008. His research interests include image/video processing, pattern recognition, image compression and content retrieval.

Shao-Zi Li received the B.S. degree from the Computer Science Department, Hunan University in 1983, and the M.S. degree from the Institute of System Engineering, Xi'an Jiaotong University in 1988, and the Ph.D. degree from the College of Computer Science, National University of Defense Technology in 2009. He currently serves as the Professor and Chair of School of Information Science and Technology of Xiamen University, the Vice Director of Fujian Key Lab of the Brain-like Intelligence System, and the Vice Director and General Secretary concurrently of the Council of Fujian Artificial Intelligence Society. His research interests cover Artificial Intelligence and Its Applications, Moving Objects Detection and Recognition, Machine Learning, Computer Vision, Natural Language Processing and Multimedia Information Retrieval, Network Multimedia and CSCW Technology and others.