Pattern Recognition ▮ (▮▮▮▮) ▮▮▮–▮▮▮

Contents lists available at ScienceDirect

Pattern Recognition

journal homepage: www.elsevier.com/locate/pr

Image retrieval based on micro-structure descriptor

Guang-Hai Liu [a,*], Zuo-Yong Li [b], Lei Zhang [c], Yong Xu [d]

[a] College of Computer Science and Information Technology, Guangxi Normal University, Guilin 541004, China
[b] Department of Computer Science, Minjiang University, Fuzhou 350108, China
[c] Department of Computing, the Hong Kong Polytechnic University, Hong Kong, China
[d] Bio-Computing Research Center, Shenzhen Graduate School, Harbin Institute of Technology, China

ARTICLE INFO

Article history:
Received 12 October 2010
Received in revised form
30 January 2011
Accepted 2 February 2011

Keywords:
Image retrieval
HSV color space
Edge orientation
Micro-structure
Micro-structure descriptor

ABSTRACT

This paper presents a simple yet efficient image retrieval approach by proposing a new image feature detector and descriptor, namely the micro-structure descriptor (MSD). The micro-structures are defined based on an edge orientation similarity, and the MSD is built based on the underlying colors in micro-structures with similar edge orientation. With micro-structures serving as a bridge, the MSD extracts features by simulating human early visual processing and it effectively integrates color, texture, shape and color layout information as a whole for image retrieval. The proposed MSD algorithm has high indexing performance and low dimensionality. Specifically, it has only 72 dimensions for full color images, and hence it is very efficient for image retrieval. The proposed method is extensively tested on Corel datasets with 15,000 natural images. The results demonstrate that it is much more efficient and effective than representative feature descriptors, such as Gabor features and multi-textons histogram, for image retrieval.

1. Introduction

Images and graphics are among the most important media formats for human communication and they provide a rich amount of information for people to understand the world. With the rapid development of digital imaging techniques and internet, more and more images are available to public. Consequently, there is an increasingly high demand for effective and efficient image indexing and retrieval methods, and image retrieval has become one of the most popular topics in the field of pattern recognition and artificial intelligence. An image retrieval system is a computer system for browsing, searching and retrieving images from a large volume of digital images. Generally speaking, there are three categories of image retrieval methods, i.e., text-based, content-based and semantic-based methods.

The origin of text-based approaches for image retrieval can be traced back to 1970s. However with the widely spread digital imaging devices, textual annotation of images becomes impractical and inefficient for image representation and retrieval. Content-based image retrieval (CBIR) was then emerging in 1980s [1]. In the past three decades, researchers have successfully developed many CBIR systems, including QIBC, MARS, Virage, Photobook, FIDS, Web Seek, Netra and SIMPLIcity. Since color, texture and shape features cannot sufficiently represent image semantics, recently semantic-based image retrieval techniques have been explored [1]. Nonetheless, due to the limitations of current artificial intelligence and related techniques, semantic-based image retrieval is still an open problem. So far, CBIR is the most important and effective image retrieval method and CBIR systems are being widely studied in both academia and industry.

It is known that human visual attention is enhanced through a process of competing interactions among neurons, which selects a few elements of attention and suppresses irrelevant materials [2]. There are close relationships between low-level visual features and human visual attention system, and hence the research on how to use visual perception mechanism for image retrieval is an important yet challenging problem. In order to extract features via simulating visual processing procedures and effectively integrate color, texture, shape features and image color layout information as a whole for image retrieval, in this paper we propose a novel feature detector and descriptor, namely micro-structures descriptors (MSD), to describe image features via micro-structures.

The micro-structures are defined by computing edge orientation similarity and the underlying colors, which can effectively represent image local features. The underlying colors refer to those colors that have similar edge orientation, and they can mimic human color perception well. With micro-structures serving as a bridge, the MSD can extract and describe color, texture and shape features simultaneously. The MSD has advantages of both statistical and structural texture description approaches.

* Corresponding author. Tel./fax: +86 0773 5811621.
E-mail addresses: liuguanghai009@163.com (G.-H. Liu), cslzhang@comp.polyu.edu.hk (L. Zhang).

0031-3203/$ - see front matter Crown Copyright © 2011 Published by Elsevier Ltd. All rights reserved.
doi:10.1016/j.patcog.2011.02.003

In addition, the MSD algorithm simulates human visual perception mechanism to some extent. Our experiments on large-scale datasets show that the MSD achieves higher retrieval precision than representative texture feature descriptors, such as Gabor feature [3] and our previous work called multi-textons histogram (MTH) [4], for image retrieval.

The rest of this paper is organized as follows. In Section 2, related works are introduced. The MSD scheme is presented in Section 3. In Section 4, the performance of the MSD in image retrieval is evaluated and compared with Gabor features and MTH on two Corel datasets. Section 5 concludes the paper.

2. Related works

Various algorithms have been designed to extract the color and texture features for image retrieval. Color histogram is invariant to orientation and scale and this makes it powerful in image classification. Hence, color histogram-based image retrieval has been extensively studied and widely used in CBIR systems for its simplicity and effectiveness. However, color histogram is difficult to characterize image spatial structures. Therefore, color descriptors have been proposed to exploit the spatial information, e.g. compact color moments, color coherence vector and color correlograms [5]. In the MPEG-7 standard, the color descriptors consist of a number of histogram descriptors, such as dominant color descriptor, color layout descriptor (CLD) and scalable color descriptor (SCD) [6]. Texture features provide an important information of the smoothness, coarseness and regularity of many real-world objects such as fruit, skin, clouds, trees, bricks and fabric, etc. [7], and texture based algorithms are also widely used in CBIR systems, including the gray level co-occurrence matrices [8], Markov random field (MRF) model [9], Gabor filtering [10], local binary pattern (LBP) [11], etc. The MPEG-7 standard adopts three texture descriptors: homogeneous texture descriptor, texture browsing descriptor and the edge histogram descriptor [10].

There are some algorithms that combine color and texture features together, such as the integrative co-occurrence matrix [12], texton co-occurrences matrix [13], multi-texton histogram (MTH) [4], color edge co-occurrence histogram (CECH) [14], color auto-correlograms [5], etc. Although computing Gabor features separately for each channel can be used as a color-texture descriptor, the computational burden of Gabor filtering is relatively big.

Apart from color and texture features, the shape features are also used in CBIR. Classical methods include moment invariants [15]·[16], Fourier transform coefficients [17]·[18], edge curvature and arc length [7]. In the MPEG-7 standard, three shape descriptors are used for object-based image retrieval: 3-D shape descriptor, region-based shape derived from Zernik moments and curvature scale space (CSS) descriptor [10].

Local image feature extraction and description have been attracting a lot of attention in recent years. Various feature descriptors have been proposed by emphasizing different image properties such as pixel intensities, color, texture and edges. Many of them are distribution-based, such as the SIFT descriptor [19], PCA-SIFT descriptor [20], GLOH descriptor [21], SURF descriptor [22], shape context [23], generalized correlograms [24] and CS-LBP algorithm [25]. These methods use histograms to represent different characteristics of image appearance or object shape without requiring segmentation. Local features have advantages that they are tolerant to certain illumination changes, perspective distortions, image transformations, and they are very robust to occlusion.

The idea of applying visual attention mechanism to image retrieval and pattern recognition has been receiving increasing attention over the past 30 years. Treisman [26] proposed a hypothesis about the role of focused attention, and the feature-integration theory of attention suggests that attention must be directed serially to each stimulus in a display whenever conjunctions of more than one separable feature are needed to characterize or distinguish the possible objects presented. The human visual system exhibits remarkable ability to detect subtle differences in textures that are generated from an aggregate of elements [27,28]. Color and texture have close relationship in terms of fundamental elements and they are considered as atoms for pre-attentive human visual perception. The term "texton" was conceptually proposed by Julesz twenty years ago [27]. Chen [29] demonstrated that the visual system is sensitive to global topological properties via three experiments on tachistoscopic perception of visual stimuli. The results indicate that an extraction of global topological properties is a basic factor in perceptual organization.

To address the fundamental question of what the primitives of visual perception are, a theory of "early topological perception" has been proposed [29,30]. Mishkin et al. proposed that the visual system is organized hierarchically into two separate cortical visual pathways, one specialized for the object vision and the other for the spatial vision [31]. Goodale and Miler [32] proposed a 'what' versus 'how' division for the primate posterior cerebral cortex. According to Goodale and Milner's framework, the role of the dorsal pathway is primarily about transforming perception into action, and the 'how' model can be considered as a generalization of the 'where' model [32]. Lindeberg developed a framework for detecting salient blob-like objects without relying on priori information [33]. Itti et al. proposed a saliency model about visual attention [34]. In the saliency model, an input image is filtered in a number of low-level visual "feature channels" at multiple spatial scales for extracting features of color, intensity and orientation. Developing computational models that describe how attention is deployed within a given scene has been an important challenge for computational neuroscience [35].

3. Micro-structure descriptor (MSD)

The contents in digital images can vary significantly so that directly comparing them is infeasible for applications such as image retrieval. However, the local structures of images from the same class (e.g. textile, mountains, etc.) often show a certain amount of similarity. The structural approach assumes that texture is formed with simple primitives called "texels" (texture elements) by following some placement rules. For example, the local binary pattern [11] can be considered as a type of texture elements. A typical example of "texels" is Julesz's textons theory [27]·[28], but it emphasizes on regular texture images. To address this problem, the concept of micro-structures is proposed in this paper for image retrieval. In some sense, we may think that the meaningful content of natural images is composed of many universal micro-structures. Therefore, if we could extract these micro-structures and describe them effectively, they can serve as common bases for the comparison and analyses of different images. This is the essential idea of this paper, and we call the proposed technique micro-structure descriptor (MSD).

One main problem of the MSD is how to define micro-structures. As an early feature-analysis approach, the feature-integration theory [26] proposed by Treisman adopts a 'two stage model'. In the pre-attentive stage, primitive features such as colors and orientation are extracted effortlessly and registered in special modules of feature maps. In the attentive stage, focal

Please cite this article as: G.-H. Liu, et al., Image retrieval based on micro-structure descriptor, Pattern Recognition (2011), doi:10.1016/j.patcog.2011.02.003

attention is required to recombine the separate features to form objects [26,30]. This two stage model of feature-integration theory inspires the proposed MSD framework.

For a full color image $g(x,y)$ of size $W \times N$, we transform the RGB color space to HSV color space for detecting the micro-structure features. In the HSV color space, we quantize the color image into 72 colors and detect the edge orientation. Then, the micro-structures are defined in the edge orientation image, and the MSD is built based on the underlying colors in micro-structures. Finally, the MSD is used to represent the image features for image retrieval. In the following, we describe these steps in detail.

3.1. HSV color space and color quantization

The HSV color space is defined in terms of three components: Hue (H), Saturation (S) and Value (V). The HSV color space can be modeled as a cylinder [7,36,37]. The H component describes the color type. It ranges 0–360°, with red at 0°, green at 120° and blue at 240°. The S component refers to the relative purity or how much the color is polluted with white color. It ranges 0–1. The lower the saturation of a color is, the more "grayness" will be presented and the more faded color will appear. The V component is used for the amount of black that is mixed with a hue, or represents the brightness of the color. It ranges 0–1.

It is well-known that color provides powerful information for image retrieval or object recognition, even in the absence of shape information. The human eye cannot perceive a large number of colors at the same time, but it is able to distinguish similar colors well. The HSV color space could mimic human color perception well. In this paper, the *HSV* color space is adopted and we uniformly quantize the color image into 72 colors. Specifically the H, S and V color channels are uniformly quantized into 8, 3 and 3 bins, respectively, so that in total $8 \times 3 \times 3 = 72$ color combinations are obtained. The quantized color image is denoted by $C(x,y)$, and $C(x,y) = w, w \in \{0,1, \ldots, 71\}$.

3.2. Edge orientation detection in HSV color space

Edge orientation has strong influence on image perception. For instance, orientation is one of the most important texton characteristics used in pre-attentive vision [7,27,28]. The orientation map in an image represents the object boundaries and texture structures, and it provides most of the semantic information in the image. Therefore, orientation detection is an important low level image processing procedure. Many existing edge detectors, such as Sobel operator, the Prewitt operator, the Robert operator, the LOG operator and the canny operator, can be used for orientation detection. However, all these edge detectors are designed for gray level images, while a color image has three color channels. If we apply the edge detector to the three channels separately, some edges caused by the spectral variations may be missed. If we convert the full color image into a gray image, and then detect the gradient magnitude and orientation, much chromatic information will be lost. In this section, we adopt the following method in HSV color space, so that the MSD can be easily constructed in a unified framework for color images.

In Cartesian space, the dot product of vectors $a(x_1,y_1,z_1)$ and $b(x_2,y_2,z_2)$ is defined as

$$a\,b = x_1x_2 + y_1y_2 + z_1z_2 \tag{1}$$

so that

$$\cos\left(\widehat{a,\,b}\right) = \frac{a\,b}{|a|\,|b|} = \frac{x_1x_2 + y_1y_2 + z_1z_2}{\sqrt{x_1^2+y_1^2+z_1^2}\sqrt{x_2^2+y_2^2+z_2^2}} \tag{2}$$

Because the HSV color space is based on the cylinder coordinate system, we transform it into Cartesian coordinate system to calculate the angle between vectors. Let (H,S,V) be a point in the cylinder coordinate system, and (H',S',V') be the transformation of (H,S,V) in Cartesian coordinate system, $H' = S \cdot \cos(H)$, $S' = S \cdot \sin(H)$ and $V' = V$. We apply Sobel operator to each of H', S' and V' channels of a color image $g(x,y)$ in Cartesian coordinate system. The reason that we use Sobel operator is that it is less sensitive to noise than other gradient operators or edge detectors, while being very efficient [7]. The gradients along x and y directions can then be denoted by two vectors $a(H_x^{'}, S_x^{'}, V_x^{'})$ and $b(H_y^{'}, S_y^{'}, V_y^{'})$, where $H_x^{'}$ denotes the gradient in H' channel along the horizontal direction, and so on. Their norm and dot product can be defined as

$$|a| = \sqrt{(H_x^{'})^2 + (S_x^{'})^2 + (V_x^{'})^2} \tag{3}$$

$$|b| = \sqrt{(H_y^{'})^2 + (S_y^{'})^2 + (V_y^{'})^2} \tag{4}$$

$$a\,b = H_x^{'}H_y^{'} + S_x^{'}S_y^{'} + V_x^{'}V_y^{'} \tag{5}$$

The angle between a and b is then

$$\cos\left(\widehat{a,\,b}\right) = \frac{a\,b}{|a|\,|b|} \tag{6}$$

$$\theta = arc\cos\left(\widehat{a,\,b}\right) = arc\cos\left[\frac{a\,b}{|a| \cdot |b|}\right] \tag{7}$$

After the edge orientation θ of each pixel is computed, the orientation is uniformly quantized into m bins, where $m \in \{6,12,18,24,30,36\}$. Denoted by $\theta(x,y)$ the edge orientation map, as $\theta(x,y) = \phi, \phi \in \{0,1, \ldots, m\}$. In Section 4.4, the experiments demonstrated that using six bins in HSV color space works well under our framework. Thus, the orientations are quantized into six bins with an interval of 30°.

3.3. Micro-structure definition and map extraction

Natural scenes are rich in color, texture and shape information, and a wide range of natural images can be considered as a mosaic of regions with different colors, textures and shapes. Edge orientation plays an important role in the human visual system for recognition and interpretation [1]. It contains rich texture and shape information. The approaches to texture description can be roughly classified into three categories: statistical, spectral and structural methods [7]. The statistical approach characterizes texture by its gray level statistics. The spectral approach is based on power spectral analysis and filtering theory in the frequency domain. The structural approach assumes that texture is formed with simple primitives called "texels" (texture elements) by following some placement rules.

A typical example of "texels" method is Julesz's texton theory [27,28]. In general, textons are defined as a set of blobs or emergent patterns sharing a common property; however, defining textons remains a challenge. Based on the texton theory, texture can be decomposed into elementary units: the texton classes of colors, elongated blobs of specific widths, orientation and aspect ratios and the terminators of these elongated blobs. In Julesz's texton theory, it does not specify details of spatial confinement of line segments and does not specify exactly the line segments. Moreover Julesz's texton theory does not consider color or natural images.

Human visual system is sensitive to orientation and color. Orientation is a powerful visual cue about the subject depicted in an image. Strong orientation usually indicates a definite pattern; however, many natural scenes do not show strong orientation and have no clear structure or specific pattern. Although the natural

Please cite this article as: G.-H. Liu, et al., Image retrieval based on micro-structure descriptor, Pattern Recognition (2011), doi:10.1016/j.patcog.2011.02.003

images show various contents, they may have some common fundamental elements. The different combination and spatial distribution of those basic elements result in the various micro-structures or patterns in the natural images. In this paper, micro-structures are defined as the collection of certain underlying colors. The underlying colors are those colors which have similar or the same edge orientation in uniform color space. The highlight of underlying colors is that they can combine color, texture and shape cues as a whole. Julesz's texton theory mainly focuses on analyzing regular textures, while the micro-structures can be considered as the extension of Julesz's textons or the color version of textons. Since micro-structures involve color, texture and shape information, they can better present image features for image retrieval.

In order to find the micro-structures, which have similar attributes such as edge orientation and color distribution, we partition the image into many small blocks, which can be a grid of size 2×2, 3×3, 5×5, 7×7 and so on. For the convenience of expression, the 3×3 block is used in the following development of micro-structure analysis. The edge orientation image $\theta(x,y)$ is used to define micro-structures, because an edge orientation is insensitive to color and illumination variation and it is independent of translation, scaling and small rotation [7]. Note that since we quantize the orientation into six levels, the values of the pixels in $\theta(x,y)$ can only vary from 0 to 5. We move the 3×3 block from left-to-right and top-to-bottom throughout the image to detect micro-structures.

In the 3×3 block, if one of the eight nearest neighbors has the same value as the center pixel, then it is kept unchanged; otherwise it is set to empty. If all the eight nearest neighboring pixels are empty, then the 3×3 block is not considered as a micro-structure and all pixels in the 3×3 block will be set to empty. The pattern resulted from this operation is called a fundamental micro-structure. Fig. 1 shows an example of the micro-structure detection process.

Suppose there is an edge orientation map $\theta(x,y)$ of size $W\times N$. When we move the 3×3 block from left-to-right and top-to-bottom throughout the image, the detected fundamental micro-structures in a neighborhood can overlap. To obtain a final single micro-structure map of the whole image, we use a simple five-step strategy described as follows.

(1) Starting from the origin (0,0), we move the 3×3 block from left-to-right and top-to-bottom throughout the edge orientation image $\theta(x,y)$ with a step-length of three pixels along both horizontal and vertical directions. Then, we will obtain a micro-structure map, denoted by $M_1(x,y)$, where $0\le x\le W-1$, $0\le y\le N-1$.
(2) Starting from the location (1,0), we move the 3×3 block from left-to-right and top-to-bottom with a step-length of three pixels along both horizontal and vertical directions. Then, a micro-structure map $M_2(x,y)$ is obtained, where $1\le x\le W-1$, $0\le y\le N-1$.
(3) Similarly starting from location (0,1), we can have the third micro-structure map $M_3(x,y)$, where $0\le x\le W-1$, $1\le y\le N-1$.
(4) Starting from location (1,1), we can have the fourth micro-structure map $M_4(x,y)$, where $1\le x\le W-1$, $1\le y\le N-1$.
(5) The final micro-structure map, denoted by $M(x,y)$, is obtained by fusing the four maps based on the following rule:

$$M(x,y)=Max\{M_1(x,y),M_2(x,y),M_3(x,y),M_4(x,y)\} \quad (8)$$

Fig. 2 uses an example to illustrate the above micro-structure map extraction process. Fig. 2(a) shows the extraction of micro-structure map $M_1(x,y)$. Maps $M_2(x,y)$, $M_3(x,y)$ and $M_4(x,y)$ can be extracted similarly. Fig. 2(b) then shows the fusion of the four maps to form the final micro-structure map $M(x,y)$.

3.4. Micro-structure image

According to the neuropsychological findings, different types of stimulus are processed separately, yet simultaneously, by different neural mechanisms before the stimulus is consciously perceived as a whole [2]. The color and orientation information is processed separately, but simultaneously. After the micro-structure map $M(x,y)$ is extracted from the edge orientation image $\theta(x,y)$, we use it as a mask to extract the underlying colors information from the quantized image $C(x,y)$. Fig. 3 illustrates this process. Fig. 3(a) shows the micro-structure map; Fig. 3(b) is the color image; by imposing the micro-structure map on the image, as shown in Fig. 3(c), finally we get the micro-structure image by preserving only the colors in the micro-structure map, as shown in Fig. 3(d). All the colors outside the map are set to empty. For the convenience of expression, we use $f(x,y)$ to denote the micro-structure image. Clearly, in the formation of micro-structure image, not only the edge features, but also the color features are exploited. That is, the micro-structure map serves as a bridge to combine the color, texture and shape features as a whole.

3.5. Micro-structure feature representation

After the micro-structure image is extracted, the next step is how to describe its features so that the different images can be compared for retrieval. However, a main problem is how to stimulate visual processing to a certain extent. There are some frameworks for understanding the ventral versus dorsal organization of the posterior cortex: cortical visual systems and visuomotor systems [38]. They are often known as the 'what/where' and "what/how' pathways. The cortical visual system hypothesis is widely accepted, but still controversial. It describes two information processing streams originating in the occipital cortex, dorsal stream and ventral stream [31]-[38]. The 'what/where' pathways are responsible for object identification (e.g. color and shape) and object location (e.g. position and motion), respectively. The 'how' pathway can be considered as a generalization of the 'where' pathway, and it transforms perception into action [32]. Those analyses provide an insight into the nature of attributes determining the relationship of image feature representation and visual attention mechanism.

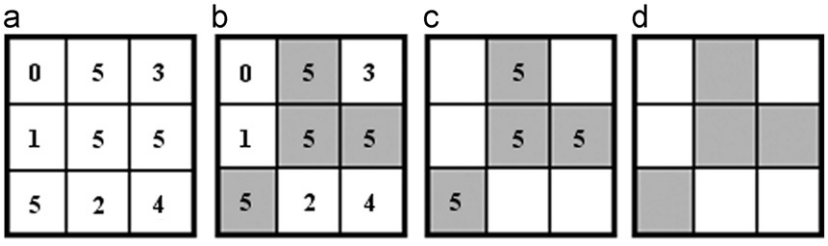

Fig. 1. An example of micro-structure detection: (a) a 3×3 grid of edge orientation map; (b)–(c) show the micro-structure detection process; and (d) shows the detected fundamental micro-structure.

Please cite this article as: G.-H. Liu, et al., Image retrieval based on micro-structure descriptor, Pattern Recognition (2011), doi:10.1016/j.patcog.2011.02.003

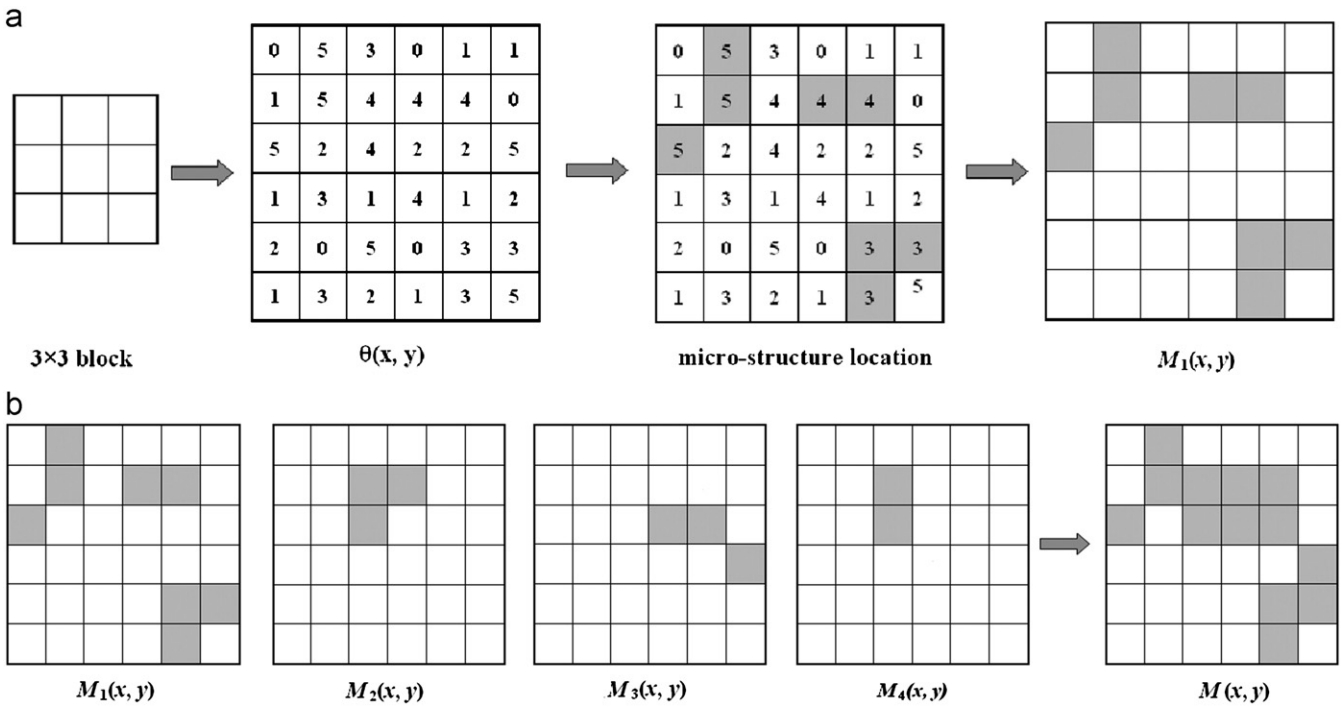

Fig. 2. Illustration of micro-structure map extraction. (a) Shows the extraction micro-structure map$M_1(x,y)$. Maps $M_2(x,y)$, $M_3(x,y)$ and $M_4(x,y)$ can be extracted similarly. (b) Shows the fusion of the four maps to form the final micro-structure map $M(x,y)$.

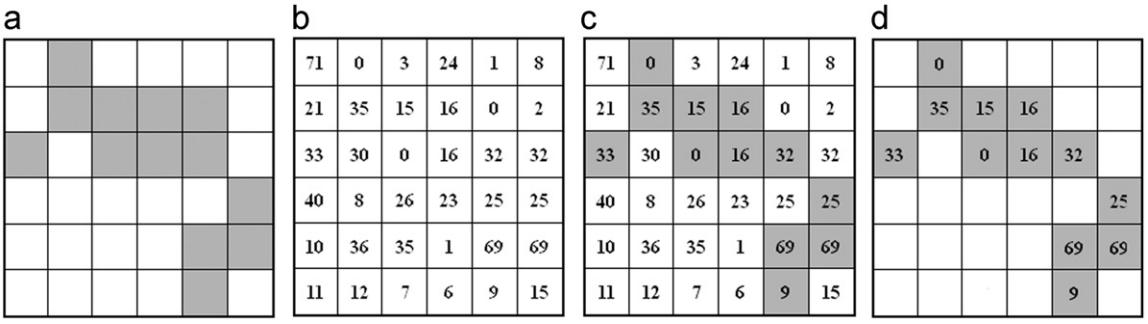

Fig. 3. Micro-structure image formation. (a) The detected micro-structure map $M(x,y)$; (b) the quantized color image $C(x,y)$; (c) imposing the micro-structure map on the image; and (d) the obtained micro-structure image $f(x,y)$ by keeping only the colors within the micro-structure map.

A typical visual scene contains many different objects, not all of which can be fully processed by the visual system. Human visual attention is enhanced through a process of competing interactions among neurons representing all of the stimuli presented in the visual field, and only a few points of attention are selected while the other irrelevant materials are suppressed [2,39]. According to this viewpoint, a behaviorally relevant object in a cluttered field is found by rapidly shifting the spotlight from one object in the scene to the next, until the sought-for object is found [2]. In the proposed algorithm, the stimuli are the similar color pixels in the micro-structure images.

In order to describe image features via simulating the visual attention mechanism to some extent, we implement the proposed algorithm according to the following rules: (1) *what* the micro-structure is; (2) *where* the micro-structure is; (3) *how* the micro-structure correlates with others. (1) and (2) are about the representation of the perception image, and (3) is about the spatial abstraction of the micro-structures in a human brain. According to these rules, the proposed algorithm can be expressed as follows.

Values of a micro-structure image $f(x,y)$ is denoted as $f(x,y)=w, w\in\{0,1,\ldots,L-1\}$. In each 3×3 block of $f(x,y)$, denoted by $P_0=(x_0,y_0)$ the center position of it and let $f(P_0)=w_0$. Denoted by $P_i=(x_i,y_i)$ the eight nearest neighbors to P_0 and let $f(P_i)=w_i$, $i=1,2,\ldots,8$. Denoted by N the co-occurring number of values w_0 and w_i, and by $\overline{N}$ the occurring number of w_0. Moving the 3×3 block from left-to-right and top-to-bottom throughout the micro-structure image, we use the following equation to describe the micro-structure features

$$H(w_0)=\begin{cases}\dfrac{N\{f(p_0)=w_0 \wedge f(p_i)=w_i \,||p_i-p_0|=1\}}{8\,\overline{N}\{f(p_0)=w_0\}} \\ \text{where } w_0=w_i,\ i\in\{1,2,\ldots,8\}\end{cases} \tag{9}$$

The dimensionality of $H(w_0)$ is 72 for color images. It can express how the spatial correlation of neighboring underlying colors distributes in the micro-structures image. Fig. 4 shows two examples of the proposed MSD.

The micro-structure image is only a part of the full color image. It fits the attribute of the human visual system, in that it selects only a few points of attention and suppresses the irrelevant material. Generally speaking, the proposed algorithm not only describes "what" and "how" colors and orientations are used, but also specifies "where" and "how" the color and

Please cite this article as: G.-H. Liu, et al., Image retrieval based on micro-structure descriptor, Pattern Recognition (2011), doi:10.1016/j.patcog.2011.02.003

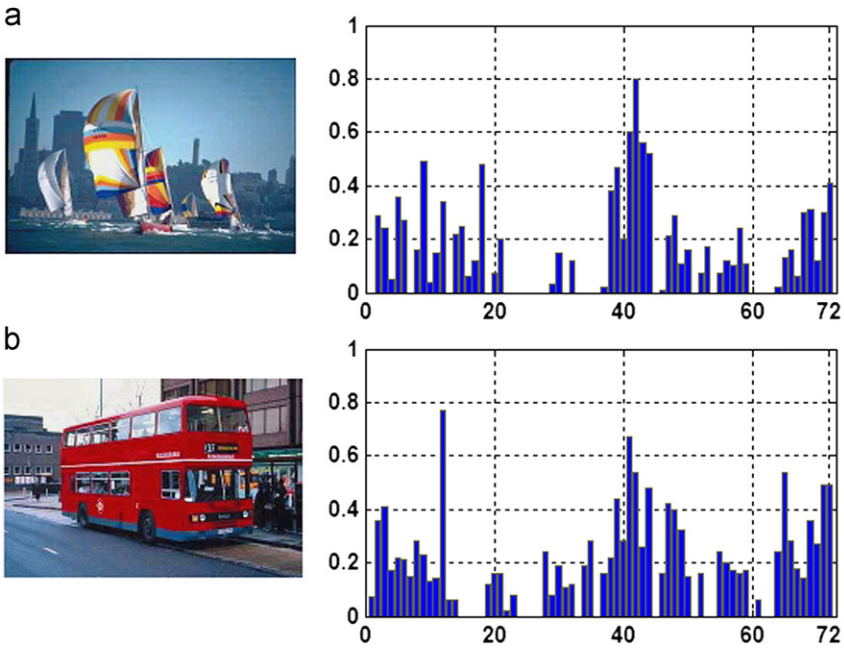

Fig. 4. Two examples of an MSD: (a) sailing ship and (b) bus.

orientation components are distributed within a certain distance in the visual scenery. The MSD can describe the different combination and spatial distribution of the micro-structures, so it has the discrimination power of color, texture, shape features and color layout information.

4. Experimental results

For a fair evaluation of MSD's performance in image retrieval, we selected those algorithms which are specifically developed for image retrieval, such as Gabor features [3] and multi-textons histogram (MTH) [4], in the experiments for comparison. In the experiments, we selected randomly 50 images from each category as query images. The performance is evaluated by the average result of each query. An online image retrieval system by the proposed method is available at: http://www.ci.gxnu.cn/cbir/.

The well-known local feature descriptors, LBP, SIFT and SURF descriptors, were originally developed for texture classification and image matching. In the SIFT and SURF descriptors, each keypoint has a 128 and 64 dimensional feature vector, respectively. This will lead to a very high dimensional descriptor for large scale image retrieval, because each image will have many keypoints. It is not suitable to use SIFT and SURF directly for image retrieval, especially for large scale image datasets. The LBP algorithm was originally developed for texture classification or analysis, and it cannot describe the smooth areas well. The CS-LBP [25] incorporates the advantage of SIFT algorithm and it can overcome this limitation. However, it is developed for object recognition, but not for image retrieval. With the above considerations, we compare MSD with Gabor features and MTH in the following experiments. The source code of our MSD algorithm can be downloaded at http://www4.comp.polyu.edu.hk/~cslzhang/code/MSD.rar

4.1. Datasets

So far there are no standard test dataset and performance evaluation model for CBIR systems [1]. Most of the researchers use Corel image dataset to test image retrieval performance, while some researchers use self-collected images or Brodatz and Outex texture datasets in experiments. However, Corel dataset has become a de-facto standard in demonstrating the performance of CBIR systems. In this section, we evaluate the performance of our method by using Corel dataset. All Corel images come from Corel Gallery Magic 20, 0000 (8 cds).

Corel image database contains a large amount of images of various contents ranging from animals and outdoor sports to natural scenarios. Two Corel datasets are used in our image retrieval systems. The first one is Corel-5000 dataset, which contains 50 categories. There are 5000 images from diverse contents such as fireworks, bark, microscopic, tile, food texture, tree, wave, pills and stained glass. Each category contains 100 images of size 192×128 in the JPEG format. The second dataset is Corel-10,000 dataset. It contains 100 categories, and there are 10,000 images from diverse contents such as sunset, beach, flower, building, car, horses, mountains, fish, food, door, etc. Each category contains 100 images of size 192×128 in the JPEG format. Corel-10,000 dataset contains all categories of Corel-5000 dataset.

4.2. Distance measure

For each template image in the dataset, an M-dimensional feature vector $T=[T_1,T_2, \ldots, T_M]$ is extracted and stored in the database. Let $Q=[Q_1,Q_2, \ldots, Q_M]$ be the feature vector of a query image, the L_1 distance between them is simply calculated as

$$D(T, Q) = \sum_{i=1}^{M} |T_i - Q_i| \tag{10}$$

The L_1 distance is simple to calculate, which needs no square or square root operations. It can save much computational cost and is very suitable for large scale image datasets. For the proposed MSD, $M=72$ for color images. The class label of the template image, which yields the smallest distance, is assigned to the query image.

Please cite this article as: G.-H. Liu, et al., Image retrieval based on micro-structure descriptor, Pattern Recognition (2011), doi:10.1016/j.patcog.2011.02.003

4.3. Performance evaluation metrics

The precision and recall indices are used to evaluate the performance of the proposed method. The two indices are the most commonly used measurements for evaluating image retrieval performance. Precision is the ratio of the number of retrieved similar images to the number of retrieved images, while recall is the ratio of the number of retrieved similar images to the total number of similar images. They are defined as follows:

$$P = I_N/N \tag{11}$$

$$R = I_N/K \tag{12}$$

where I_N is the number of similar images retrieved, N is the total number of images retrieved and K is the total number of similar images. In our image retrieval system, we set $N=12$ and $K=100$ for Corel datasets.

4.4. Retrieval results

We use the RGB, HSV and Lab color spaces to evaluate the retrieval performance. The average retrieval precisions and recalls are listed in Tables 1–3. According to the retrieval results, the proposed algorithm achieves the best retrieval performance in the HSV color space. The retrieval precision may decrease when the number of quantization levels is too high. And a too fine quantization level cannot suppress the noise and variations in the image. On the other hand, a finer quantization level will lead to a higher dimensionality of the feature vector, so that more storage space and more search time are required. However, a too coarse quantization reduces the discrimination power. For a good balance between retrieval accuracy, storage space and retrieval speed, we set the color and orientation quantization level in the MSD as 72 and 6 for image retrieval. The proposed algorithm performs poorly in RGB color space, because RGB color space is not perceptually uniform and does not separate the luminance component from the chrominance ones, and hence affect the retrieval performance of the proposed algorithm.

The HSV and Lab color spaces are perceptually uniform. The H component is particularly important in the HSV color space, since it represents color in a manner that mimics human color recognition. So we set the uniform quantization number of H component $bin(H) \geq 8$, and other two components $bin(S) \geq 3$ and $bin(V) \geq 3$, in the proposed framework. As can be seen from the results in Tables 1–3, the proposed MSD works much better in these two perceptually uniform color spaces than in the RGB color space. In addition, the HSV color space has slightly better performance than the Lab color space. Thus, the HSV color space is employed in the proposed algorithm.

Table 4 lists the average retrieval precision and recall by the MSD with different distances or similarity metrics. As can be seen

Table 1
The average retrieval precision and recall of the MSD under different color and orientation quantization levels on Corel-5000 dataset in RGB color space.

Color quantization levels	Texture orientation quantization levels											
	Precision (%)						Recall (%)					
	6	12	18	24	30	36	6	12	18	24	30	36
128	50.46	50.24	50.03	49.71	49.79	50.01	6.06	6.03	6.00	5.97	5.98	6.00
64	49.42	49.32	48.77	48.87	48.89	49.14	5.93	5.92	5.85	5.86	5.87	5.90
32	46.70	46.89	46.31	46.28	46.51	46.52	5.61	5.63	5.56	5.56	5.58	5.58
16	39.08	39.53	39.31	39.47	39.18	39.26	4.69	4.75	4.72	4.74	4.70	4.71

Table 2
The average retrieval precision and recall of the MSD under different color and orientation quantization levels on Corel-5000 dataset in HSV color space.

Color quantization levels	Texture orientation quantization levels											
	Precision (%)						Recall (%)					
	6	12	18	24	30	36	6	12	18	24	30	36
192	58.04	57.9	57.74	57.45	57.68	57.56	6.96	6.95	6.93	6.89	6.92	6.91
128	58.06	57.58	57.84	57.74	57.59	57.49	6.97	6.91	6.94	6.93	6.91	6.90
108	56.89	56.54	56.45	56.73	56.72	56.17	6.83	6.78	6.77	6.81	6.81	6.74
72	55.92	55.95	56.02	56.15	56.5	55.79	6.71	6.71	6.72	6.74	6.78	6.70

Table 3
The average retrieval precision and recall of the MSD under different color and orientation quantization levels on Corel -5000 dataset in Lab color space.

Color quantization levels	Texture orientation quantization levels											
	Precision (%)						Recall (%)					
	6	12	18	24	30	36	6	12	18	24	30	36
225	54.34	53.71	53.64	53.50	53.09	53.36	6.52	6.45	6.44	6.42	6.37	6.40
180	54.87	54.40	54.70	54.51	54.26	54.42	6.59	6.53	6.56	6.54	6.51	6.53
90	53.17	53.40	53.46	53.07	53.34	53.30	6.38	6.41	6.42	6.37	6.40	6.40
45	47.16	46.84	47.58	46.97	47.41	47.36	5.66	5.62	5.71	5.64	5.69	5.69

Please cite this article as: G.-H. Liu, et al., Image retrieval based on micro-structure descriptor, Pattern Recognition (2011), doi:10.1016/j.patcog.2011.02.003

Table 4
The average retrieval precision and recall by the MSD with different distances or similarity metrics.

Dataset	Performance	Distance or similarity metrics						
		Canberra	L_1	L_2	Quadratic	Weighted L_1	Cos correlation	Histogram intersection
Corel-5000	Precision (%)	49.73	55.92	55.92	24.76	55.25	54.08	34.41
	Recall (%)	5.97	6.71	6.71	2.98	6.63	6.49	4.13
Corel-10,000	Precision (%)	40.05	45.62	45.62	19.01	45.14	43.49	23.83
	Recall (%)	4.81	5.48	5.48	2.29	5.42	5.22	2.87

Table 5
The average retrieval precision and recall ratios by various methods on Corel datasets.

Dataset	Precision (%)			Recall (%)		
	Gabor	MTH	MSD	Gabor	MTH	MSD
Corel-5000	36.22	49.84	55.92	4.35	5.98	6.71
Corel-10000	29.15	41.44	45.62	3.50	4.97	5.48

from it, among these distance or similarity metrics, L_1 and L_2 metrics perform the best and surpass other metrics, and the quadratic metric gives the worst results. In the proposed framework, we adopted L_1 metric, but not L_2 metric because the L_1 distance is very simple to calculate and needs no square or square root operations. It can save much computational cost and is very suitable for large scale image datasets. The Canberra distance is also simple to calculate and it obtains good precision in retrieval experiments. As can be seen in Fig. 4, there are some bins whose frequencies are close to zero. Thus, if we apply the histogram intersection to the MSD, the probability that $min(T,Q)=0$ will be high and some false matches may appear. Therefore, the histogram intersection is also not suitable to the proposed MSD as a similarity metric.

L_1, L_2 and histogram intersection metrics belong to Bin-by-Bin type; Canberra, weighted L_1 and Cos correlation metrics belong to the weighted type; and quadratic metric belongs to the Cross-Bin type. As can be seen from Table 4, the L_1 metric performs the best and surpasses other Bin-by-Bin, Cross-Bin and weighted L_1 metrics (in weighted L_1 metric, $1/(1+T_i+Q_i)$ is the weight). Generally speaking, the Cross-Bin metric can obtain better performance than that of Bin-by-Bin metric and histogram intersection metric for color histogram-based image retrieval. Indeed, the proposed algorithm is not a traditional color histogram, but a specific histogram that combines color, orientation and spatial layout information as a whole, so the quadratic form metric is not suitable for this framework. However, the computation burden of the cross-bin type metric is also bigger than that of the bin-by-bin type metrics. So we adopt the L_1 metric in Eq. (10) for matching.

The precision vs. recall results by Gabor feature, MTH and MSD methods on the two Corel dataset are listed in Table 5. It can be seen that the proposed MSD achieves much better performance than the other methods. On Corel-5000 dataset, it outperforms Gabor feature and MTH by 19.70% and 6.08%, respectively. On Corel-10,000 dataset, MSD outperforms Gabor feature and MTH by 16.47% and 4.48%, respectively, in terms of the average retrieval precision.

To better illustrate the retrieval effectiveness of our algorithm, we plot the precision-recall curves in Fig. 5. The horizontal axis corresponds to the recall, while the vertical axis corresponds to precision. In image retrieval, if a method achieves higher precision and recall for large answer sets, it is considered to be better than others; if the average retrieval precision and recall is higher, curves will go far from the origin of coordinate; if the performance of two methods on some image categories are very similar, part of their curves may be overlapped. In general, if the performance of a method is low, its curve will be short and concentrated on a certain field. If a method has reasonable performance over each image category, its curve will be smooth. If there are many significant turning points on curves, it means that the performance is unstable. As can be seen from Fig. 5, MSD performs much better than Gabor feature and MTH methods on the two Corel datasets. Its precision vs. recall curve is far from the origin of the coordinate. The number of turning points is few and curve spans well for all image categories.

Figs. 6 and 7 show two examples of the image retrieval on Corel-10,000 dataset by the proposed algorithm. As can be seen from Fig. 6, on Corel-10000dataset, the retrieved images by MSD have very similar color and texture features or scene contents. From Fig. 7, we see that an MSD can describe color and shape features well. Most of the retrieved images have very similar color appearance to the query image. These examples validate that the MSD can capture the common micro-structures of natural images, and it can represent the distribution of edge orientation and color features via micro-structures.

Many texture descriptors can obtain good performance only in regular texture images. However, the performance of these texture descriptors, such as EHD [6], Gabor feature [3,6,10,40], etc., will be reduced when processing natural images. Gabor filter is widely used to extract texture features for image retrieval, because frequency and orientation representations of Gabor filter are similar to those of the human visual system, and it has been found to be particularly appropriate for texture representation and discrimination [40]. However, as images in real-world often have no homogenous textures or regular textures, the image features obtained from Gabor filtering cannot represent the property of real-world images well. On the other hand, the texture feature only represents partial attribute of images, and using a single attribute to describe image features is not accurate enough for image retrieval.

MTH [4] combines the first- and second-order statistics into an entity for the texton analysis, and thus the texture discrimination power is greatly increased. The MTH can represent the spatial correlation of edge orientation and color based on textons analysis [4]. Its performance is better than that of Gabor filter. The MTH analyzes the spatial correlation between neighboring colors and edge orientations based on four special texton types, while these four special texton types are only a part of many texton types in the natural image. This limits the discrimination power of the MTH, and makes the MTH hard to fully represent the content of texton images.

The proposed MSD algorithm analyzes the spatial correlation among neighboring micro-structures by the uniform color space. The micro-structure can efficiently combine color, texture and shape cue as a whole, so it can be considered as an extension of

Please cite this article as: G.-H. Liu, et al., Image retrieval based on micro-structure descriptor, Pattern Recognition (2011), doi:10.1016/j.patcog.2011.02.003

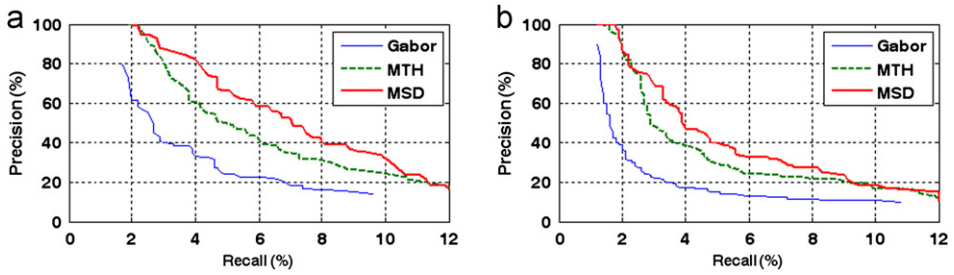

Fig. 5. The precision vs. recall curves by Gabor, MTH and MSD: (a) Corel-5000 dataset and (b) Corel-10,000 dataset.

Fig. 6. An example of image retrieval by the MSD on Corel −10,000 dataset. The query is a sailing ship image, and 9 images are correctly retrieved and ranked within the top 12 images. (The top-left image is the query image, and the similar images include the query image itself.)

Julesz's texton or the color version of texton. The MSD also overcomes the drawback of the MTH by introducing micro-structures and simulating human brain's working procedures for visual information processing. The MSD can describe the different combination and spatial distribution of micro-structures, and has the discrimination power of color, texture, shape features and layout information. At the same time, the vector dimension in the proposed algorithm is smaller than that of the MTH algorithm.

All experiments were conducted on an Intel (R) Core (TM) 2 Quad 2.83 GHz PC with 4 GB memory and the Windows XP operating system. The image retrieval system is built in Visual C# 2010. During the course of features extraction for a natural image of size 192×128, the average time usage of Gabor filter, MTH and MSD are 1326.33, 71.18 and 54.64 ms, respectively. The computational burden of implementing Gabor filter is the highest, because it needs to perform filtering along various scales and octaves. The time used by the MSD is mainly on the stage of micro-structure analysis and description. According to the computation burdens, it is clear that the proposed algorithm is very suitable for image retrieval on large scale image datasets.

5. Conclusion

A simple yet efficient image retrieval approach, namely micro-structure descriptor (MSD), was developed in this paper. The micro-structures are defined by an edge orientation similarity with the underlying colors, which can effectively represent image features. The underlying colors are colors with similar edge orientation and can mimic the human color perception well. With micro-structures serving as a bridge, the MSD can extract and describe color, texture and shape features simultaneously. The MSD has advantages of both statistical and structural texture description approaches. In addition, this algorithm simulates human visual perception mechanism to a certain extent. The MSD algorithm has higher indexing performance and efficiency for image retrieval, but lower dimensionality which is only 72 for

Please cite this article as: G.-H. Liu, et al., Image retrieval based on micro-structure descriptor, Pattern Recognition (2011), doi:10.1016/j.patcog.2011.02.003

Fig. 7. An example of image retrieval by the MSD on Corel −10,000 dataset. The query is a bus image, and 11 images are correctly retrieved and ranked within the top 12 images. (The top-left image is the query image, and similar images include the query image itself.)

full color images. Our experiments on large-scale datasets show that the MSD achieves higher retrieval precision than the existing representative image feature descriptors, such as Gabor feature and MTH, for image retrieval. It has good discrimination power of color, texture, shape features and layout information.

Acknowledgment

This work was supported by the Hong Kong RGC General Research Fund (PolyU 5351/08E) and the National Natural Science Fund of China (No. 60632050, No. 60775015). The authors would like to thank the anonymous reviewers for their constructive comments.

References

[1] Y. Liu, D. Zhang, G. Lu, W.Y. Ma, A survey of content-based image retrieval with high-level semantics, Pattern Recognition 40 (11) (2007) 262–282.
[2] R. Desimone, Visual attention mediated by biased competition in extrastriate visual cortex, Philosophical Transactions of the Royal Society B 353 (1998) 1245–1255.
[3] B.S. Manjunath, W.Y. Ma, Texture features for browsing and retrieval of image data, IEEE Transactions on Pattern Analysis and Machine Intelligence 18 (8) (1996) 837–842.
[4] G-H. Liu, L. Zhang, et al., Image retrieval based on multi-texton histogram, Pattern Recognition 43 (7) (2010) 2380–2389.
[5] J. Huang, S.R. Kumar, M. Mitra, et al., Image indexing using color correlograms, in: IEEE Conference on Computer Vision and Pattern Recognition, (1997) 762–768.
[6] B.S. Manjunath, J-R. Ohm, et al., Color and texture descriptors, IEEE Transactions on Circuit and Systems for Video Technology 11 (6) (2001) 703–715.
[7] R.C. Gonzalez, R.E. Woods, Digital Image Processing, third ed., Prentice Hall, 2007.
[8] R.M. Haralick, k. Shangmugam, L. Dinstein, Textural feature for image classification, IEEE Transactions on System, Man and Cybernetics SMC-3 (6) (1973) 610–621.
[9] G. Cross, A. Jain, Markov random field texture models, IEEE Transactions on Pattern Analysis and Machine Intelligence 5 (1) (1983) 25–39.
[10] B.S. Manjunath, P. Salembier, T. Sikora, Introduction to MPEG-7: Multimedia Content Description Interface, John Wiley & Sons Ltd., 2002.
[11] T. Ojala, M. Pietikanen, T. Maenpaa, Multi-resolution gray-scale and rotation invariant texture classification with local binary patterns, IEEE Transactions on Pattern Analysis and Machine Intelligence 24 (7) (2002) 971–987.
[12] C. Palm, Color texture classification by integrative co-occurrence matrices, Pattern Recognition 37 (5) (2004) 965–976.
[13] G-H. Liu, J-Y. Yang, Image retrieval based on the texton co-occurrence matrix, Pattern Recognition 41 (12) (2008) 3521–3527.
[14] J. Luo, D. Crandall, Color object detection using spatial-color joint probability functions, IEEE Transactions on Image Processing 15 (6) (2006) 1443–1453.
[15] M-K. Hu, Visual pattern recognition by moment invariants, IRE Transactions on Information Theory (1962) 179–187.
[16] P-T. Yap, R. Paramesran, S-H. Ong, Image analysis using Hahn moments, IEEE Transactions on Pattern Analysis and Machine Intelligence 29 (11) (2007) 2057–2062.
[17] F.P. Kuhl, C.R. Giardina, Elliptic Fourier features of a closed contour, Computer Graphics Image Process 18 (1982) 236–258.
[18] C.T. Zahn, R.Z. Roskies, Fourier descriptors for plane closed curves, IEEE Transactions on Computer C-21 3 (1972) 269–281.
[19] D.G. Lowe, Distinctive image features from scale-invariant keypoints, International Journal of Computer Vision 60 (2) (2004) 91–110.
[20] Y. Ke, R. Sukthankar, PCA-SIFT: a more distinctive representation for local image descriptors, in: IEEE Conference on Computer Vision and Pattern Recognition, (2004) 506–513.
[21] K. Mikolajczyk, C. Schmid, A performance evaluation of local descriptors, IEEE Transactions on Pattern Analysis and Machine Intelligence 27 (10) (2005) 1615–1630.
[22] H. Bay, T. Tuytelaars, L.V. Gool, SURF: speeded up robust features, in: European Conference on Computer Vision, (2006) 404–417.
[23] S. Belongie, J. Malik, J. Puzicha, Shape matching and object recognition using shape contexts, IEEE Transactions on Pattern Analysis and Machine Intelligence 24 (4) (2002) 509–522.
[24] J. Amores, N. Sebe, P. Radeva, Context based object-class recognition and retrieval by generalized correlograms, IEEE Transactions on Pattern Analysis and Machine Intelligence 29 (10) (2007) 1818–1833.
[25] M. Heikkila, M. Pietiaine, C. Schmid, Description of interest regions with local binary patterns, Pattern Recognition 42 (3) (2009) 425–436.
[26] A. Treisman, A feature in integration theory of attention, Cognitive Psychology 12 (1) (1980) 97–136.
[27] B. Julesz, Textons, the elements of texture perception and their interactions, Nature 290 (5802) (1981) 91–97.
[28] B. Julesz, Texton gradients: the texton theory revisited, Biological Cybernetics 54 (1986) 245–251.
[29] L. Chen, Topological structure in visual perception, Science 218 (4573) (1982) 699–700.
[30] L. Chen, The topological approach to perceptual organization, Visual Cognition 12 (4) (2005) 553–637.
[31] M. Mishkin, L.G. ungerleider, et al., Object vision and spatial vision—two cortical pathways, Trends in Neuroscience 6 (1983) 414–417.
[32] M.A. Goodale, A.D. Milner, Separate visual pathways for perception and action, Trends in Neurosciences 15 (1) (1992) 20–25.

Please cite this article as: G.-H. Liu, et al., Image retrieval based on micro-structure descriptor, Pattern Recognition (2011), doi:10.1016/j.patcog.2011.02.003

[33] T. Lindeberg, Detecting salient blob-like image structures with a scale-space primal sketch: a method for focus-of-attention, International Journal of Computer Vision 11 (3) (1993) 283–318.
[34] L. Itti, C. Koch, E. Niebur, A model of saliency-based visual attention for rapid scene analysis, IEEE Transactions on Pattern Analysis and Machine Intelligence 20 (11) (1998) 1254–1259.
[35] L. itti, C. Koch, Computational modeling of visual attention, Nature Reviews Neuroscience 2 (3) (2001) 94–203.
[36] P.W.M Tsang, W.H. Tsang, Edge detection on object color, in: Proceedings International Conference on Image Processing, (1996) 1049–1052.
[37] W.H. Tsang, P.W.M. Tsang, Edge gradient method on object color, IEEEE TENCON-Digital Signal Processing Application (1996) 310–349.
[38] R.c. O'Reily, The what and how of prefrontal cortical organization, Trends in Neurosciences 33 (8) (2010) 355–361.
[39] A.A. Rensink, J.K. O'Regan, J.J. Clark, To see or not to see: the need for attention to perceive changes in scenes, Psychological science 8 (5) (1997) 368–373.
[40] M. Kokare, P.K. Biswas, B.N. Chatterji, Texture image retrieval using rotated wavelet filters, Pattern Recognition Letters 28 (10) (2007) 1240–1249.

Guang-Hai Liu is currently an Associate Professor in the College of Computer Science and Information Technology, Guangxi Normal University in China. He received Ph.D. degree from the School of Computer Science and Technology, Nanjing University of Science and Technology (NUST). His current research interests are in the areas of image processing, pattern recognition and artificial intelligence.

Zuo-Yong Li received the B.S. degree in computer science and technology from the Fuzhou University in 2002. He got his M.S. degree in Computer Science and Technology from the Fuzhou University in 2006. Now, he is a Ph.D. Candidate in the Nanjing University of Science and Technology and a lecturer in the Department of Computer Science of the Minjiang University. He has published several papers in international/national journals. His research interests include image segmentation and pattern recognition.

Lei Zhang received the B.S. degree in 1995 from the Shenyang Institute of Aeronautical Engineering, Shenyang, PR China, the M.S. and Ph.D. degrees in Electrical and Engineering from the Northwestern Polytechnic University, Xi'an, PR China, respectively, in 1998 and 2001. From 2001 to 2002, he was a research associate in the Department of Computing, The Hong Kong Polytechnic University. From January 2003 to 2006 he worked as a Postdoctoral Fellow in the Department of Electrical and Computer Engineering, McMaster University, Canada. Since January 2006, he has been an Assistant Professor in the Department of Computing, The Hong Kong Polytechnic University. His research interests include Image and Video Processing, Biometrics, Pattern Recognition, Multi-sensor Data Fusion and Optimal Estimation Theory, etc.

Yong Xu received his B.S. and M.S. degrees at Air Force Institute of Meteorology (China) in 1994 and 1997, respectively. He then received his Ph.D. degree in pattern recognition and intelligence system at the Nanjing University of Science and Technology (NUST) in 2005. From May 2005 to April 2007, he worked at Shenzhen graduate school, Harbin Institute of Technology (HIT) as a postdoctoral research fellow. Now he is an associate professor at Shenzhen graduate school, HIT. He also acts as a research assistant researcher at the HongKong Polytechnic University from August 2007 to June 2008. His current interests include pattern recognition, biometrics, and machine learning. He has published more than 40 scientific papers.

Please cite this article as: G.-H. Liu, et al., Image retrieval based on micro-structure descriptor, Pattern Recognition (2011), doi:10.1016/j.patcog.2011.02.003