Hindawi Publishing Corporation
Mathematical Problems in Engineering
Volume 2016, Article ID 3908056, 12 pages
http://dx.doi.org/10.1155/2016/3908056

Research Article

Features Conduction Neural Response and Its Application in Content-Based Image Retrieval

Zhengfa Hu,[1] Tian Yue,[1] and Haixia Xiao[1,2]

[1]*Department of Sciences, Hubei University of Automotive Technology, Shiyan, Hubei 442002, China*
[2]*School of Automation, Huazhong University of Science and Technology, Wuhan, Hubei 430074, China*

Correspondence should be addressed to Tian Yue; yuetian@cumt.edu.cn

Received 20 February 2016; Revised 8 August 2016; Accepted 24 August 2016

Academic Editor: Simone Bianco

A novel image representation is proposed for content-based image retrieval (CBIR). The core idea of the proposed method is to do deep learning for the local features of image and to melt semantic component into the representation through a hierarchical architecture which is built to simulate human visual perception system, and then a new image descriptor of features conduction neural response (FCNR) is constructed. Compared with the classical neural response (NR), FCNR has lower computational complexity and is more suitable for CBIR tasks. The results of experiments on a commonly used image database demonstrate that, compared with those of NR related methods or some other image descriptors that were originally developed for CBIR, the proposed method has wonderful performance on retrieval efficiency and effectiveness.

1. Introduction

Driven by the demand of search service market, the method of content-based image retrieval (CBIR) becomes a hot issue in the research field of pattern recognition and artificial intelligence for many years. The common ground for CBIR systems is to extract a signature for every image based on its pixel values and to define a rule for comparing images. The components of the signature are called features. An obvious advantage of a signature over the original pixel values is the significant compression of image representation. However, a more important reason for using the signature is to gain an improved correlation between image representation and image semantics. Actually, the main task of designing a signature is to bridge the gap between image semantics and the pixel representation, that is, to create a better correlation with image semantics [1].

The researchers have tried to use machine learning techniques to derive the similarity measure of the high-level semantics of the image from the existing image representations [2] or to cluster the images by self-organizing maps firstly and then to do retrieval [3] with the former such as bandletized regions through support vector machines (BRSVM) learning and online multiple kernel similarity (OMKS) learning [4–6] and the latter such as tree structured self-organizing maps (TS-SOM) and growing hierarchical quadtree self-organizing map (GHSOQM) [7, 8]. These methods are often used in combination with relevance feedback technology, which can enhance the retrieval effectiveness to a certain extent [8, 9]. However, these methods are very technical and often need a lot of training time which makes them difficult to be applied in practice.

On the other hand, the research of image representation for CBIR is constantly advancing, and many creative image representation methods are proposed. These methods can be broadly divided into two categories: the global feature based approach and the local feature based approach. For example, the edge histogram descriptor (EHD) [10], multiple texture histogram (MTH) [11], and color difference histogram (CDH) [12] are all based on the global characteristics of the algorithm. These algorithms to extract characteristics have good identification ability and robustness. However, we know that the overly complex feature representation is not always applicable to CBIR [1, 13]. At the same time, the local feature extraction method has also been a great concern [14, 15]. These methods focus on the feature representation of

the image using the key points [16] or significant blocks in the image [17, 18]. How to determine the key points and the salient regions of the image are often dependent on the complex image segmentation technology. So far, however, the image segmentation technology is still one of the difficult problems in image processing and thus limits the application of these methods in CBIR.

In recent years, the human visual cortex neural science and the related hierarchical learning methods provide a new direction for studying of this problem. Research has shown that the human visual perception system has very good abilities of learning and generalization through a few examples, and these abilities are given by the hierarchical structure of the visual cortex [19–21]. Based on the hierarchical structure of visual cortex, Smale et al. proposed the concept of derived kernel and the related theory of neural responses (NR) [22]. They established a mathematical model to simulate the process of hierarchical processing information of the human visual system. In the NR model, the inner product defined by the neural response led to a similarity measure between images which was called the derived kernel. Based on a hierarchical architecture, a recursive definition of the neural response and associated derived kernel was given. The derived kernel can be used in a variety of application domains such as classification of images, strings of text, and genomics data. Theoretical analysis and experimental results show that the NR model is an effective feature extraction method. It has the potential to be further improved and enhanced in many applications [17, 23–25]. Most important of all, the NR model has a key semantic component: a system of templates which can fuse the visual features and the semantic features of an image together and which is very important in CBIR.

However, because of the underlying neural response using the pixel value of the bottom subblock of image and then being passed to the upper level of the subblock, this algorithm is not suitable for CBIR. Because, in the task of CBIR, the image databases are usually very large and the resolution of the image is usually very high, the exhaustion algorithm of pixel to pixel is difficult to bridge the "semantic gap" in complex scene images and a huge amount of computation also limits its application in practice. In order to capture the high-level semantic feature of the image and at the same time improve the efficiency of retrieval, we propose the concept and the corresponding algorithm of features conduction neural response (FCNR) on the basis of the related theory of NR.

In the proposed method, we divide the spatial domain of an image in a simple way firstly and then obtain the local feature representation of the image by extracting the basic characteristics such as color, texture, and shape feature on the local area of the image. Next, we establish a hierarchical structure for the local feature representation of the image; at the same time, for each layer of the structure, a local feature template set is constructed. In the first layer of the hierarchical structure, local features are used to construct initial neural response and then these features are conducted to the senior subblocks layer by layer by the normalized inner product of the neural response. Finally, the image is expressed as a vector which is called FCNR, which can be used as an image representation for CBIR. The major advantages of the proposed method can be summarized as given here.

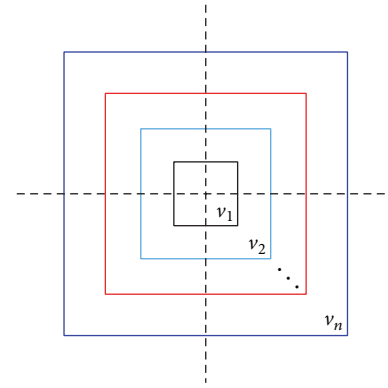

Figure 1: The n layers of nested architectures.

(i) The FCNR is derived from a local feature array rather than just using the pixel values, which overcome the drawback of overlearning problem in classical NR method.

(ii) The high-level semantic component of the image is introduced into the feature representation by the interaction between the subblocks of the image and the templates in every layer of the constructed hierarchical structure.

(iii) Without loss of the excellent identification and the invariance, the FCNR gets rid of the plight of the pixel to pixel exhaustion algorithm of the NR and reduces the computational complexity significantly, which is essential for the CBIR purpose.

The rest of this paper is organized as follows. In Section 2 the models of FCNR are constructed firstly. In Section 3 the image retrieval method based on FCNR is introduced. Then, in Section 4, we verify the effectiveness of the proposed method with extensive experiments on popular data sets and compare it with other CBIR methods. Finally, conclusions are drawn, and some future research issues are discussed in Section 5.

2. Feature Conduction Neural Response (FCNR)

The starting point of NR was to establish the mathematical model for visual mechanism of primate visual cortex [19, 26, 27]. In order to simulate the hierarchical information processing of visual cortex, Smale et al. [22] divided the image domain into some nested blocks as shown in Figure 1.

The NR of an image was defined in bottom-up fashion based on the hierarchical architecture. As a feature vector of an image, NR can be used to define the similarity between images. The theoretical analysis and experimental results show that the NR model has good performance on discrimination and it was robust to transformations, which suggested that the learning process of NR model possessed

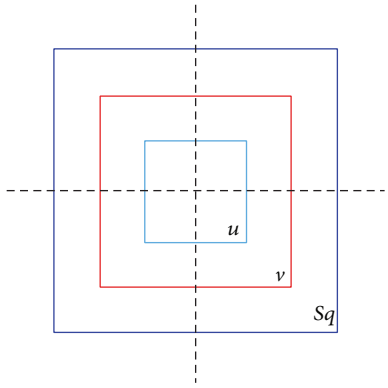

Figure 2: The three layers of nested architectures.

the characteristics of the human visual system in a certain degree.

2.1. Notation and Preliminaries. In this paper, we consider the case of a three-layer hierarchical architecture.

As shown in Figure 2, let regions u, v, and Sq in $\mathbb{R}^2$ ($u \subset v \subset Sq$) be pieces of the domain on which the patches or subpatches of images are defined. In the vision interpretation, these regions can be considered as receptive fields with different sizes. When we are working with gray scale images, an image or an image patch can be seen as a discrete function of two variables which take the corresponding gray values as the functional values. That is to say, an image of size Sq can be seen as a function defined on the domain Sq. In this case, an image set consisting of the images defined on the domain Sq can be denoted by $\mathcal{I}_{Sq}$. For description convenience, we denote the cardinality of a set A as $|A|$ in the rest of this paper. Accordingly, the images in $\mathcal{I}_{Sq}$ can be denoted by $\mathbf{f}_1, \mathbf{f}_2, \ldots, \mathbf{f}_{|\mathcal{I}_{Sq}|}$; that is, $\mathcal{I}_{Sq} = \{\mathbf{f}_1, \mathbf{f}_2, \ldots, \mathbf{f}_{|\mathcal{I}_{Sq}|}\}$. Similarly, the sets of image patches of size u and size v can be denoted by $\mathcal{I}_u$ and $\mathcal{I}_v$ with $\mathcal{I}_u = \{\mathbf{f}_1^u, \mathbf{f}_2^u, \ldots, \mathbf{f}_{|\mathcal{I}_u|}^u\}$ and $\mathcal{I}_v = \{\mathbf{f}_1^v, \mathbf{f}_2^v, \ldots, \mathbf{f}_{|\mathcal{I}_v|}^v\}$, respectively.

As an example, Figure 3 shows the nested architectures and the relationship of the image and the image patches. In Figure 3, $\mathbf{f}$ is a whole image of size 384×256, $\mathbf{f}_i^v$ ($i = 1, 2$) are two examples of image patches of size 128×128 cut out from image $\mathbf{f}$, and $\mathbf{f}_j^u$ ($j = 1, 2, 3, 4$) are four examples of image patches of size 64×64 cut out from image $\mathbf{f}_i^v$, respectively.

The elements h_i^v ($i = 1, 2$) and h_j^u ($j = 1, 2, 3, 4$) in Figure 3 which restrict an image or image patch to a specific subpatch are the other key ingredients in NR model: transformations associating two adjacent domains. Formally, the set $\mathcal{H}_u = \{h_1^u, h_2^u, \ldots, h_{|\mathcal{H}_u|}^u\}$ is called transformation set, in which the map $h^u : u \to v$ is a transformation from the smallest patch u to the next larger patch v. Similarly $\mathcal{H}_v = \{h_1^v, h_2^v, \ldots, h_{|\mathcal{H}_v|}^v\}$ with $h^v : v \to Sq$ can be defined. In this paper, the transformations are limited to translations and take the form $h(x) = x + a$. Consequently, we can consider $\mathcal{H}_u$ as a set of translations corresponding to moving a sliding window of size u in patch v and similarly $\mathcal{H}_v$ as a set of translations corresponding to moving a sliding window of size v in patch Sq. For example, given an image of size $M \times N$, if the step length equals one pixel, $(M - m + 1) \times (N - n + 1)$ image patches can be obtained by restricting the image on the given subpatch of size $m \times n$.

The following fundamental assumption related to image sets and transformation sets is supposed to be satisfied throughout this paper [22].

Axiom 1. If $\mathbf{f}^v \in \mathcal{I}_v$ and $h^u \in \mathcal{H}_u$, then $\mathbf{f}^v \circ h^u \in \mathcal{I}_u$, where $\mathbf{f}^v \circ h^u$ denotes the restriction of image patch $\mathbf{f}^v$ on region u by transformation of h^u. Similarly, $\mathbf{f} \circ h^v \in \mathcal{I}_v$ if $\mathbf{f} \in \mathcal{I}_{Sq}$ and $h^v \in \mathcal{H}_v$.

The last essential factor in NR model is series templates sets. The finite elements $\mathbf{T}^u \in \mathcal{I}_u$ are selected as the first-layer templates and the first-layer template set $\mathcal{T}_u = \{\mathbf{T}_1^u, \mathbf{T}_2^u, \ldots, \mathbf{T}_{|\mathcal{T}_u|}^u\}$ is constituted. In the same way, the second-layer template set $\mathcal{T}_v = \{\mathbf{T}_1^v, \mathbf{T}_2^v, \ldots, \mathbf{T}_{|\mathcal{T}_v|}^v\}$ can be obtained. Obviously, templates are some image patches which can be seen as image elements frequently encountered and serve as building blocks to represent other images. Those templates implicate abundant higher semantic information of the images which can be used to promote identification ability in image retrieval.

2.2. The Construction of FCNR. The first step of constructing FCNR is to segment the whole image in a simply way, which is different from other feature extraction methods based on region segmentation technology [8, 14], here just using the perpendicular line network to segment the image into some small rectangular area of the same size. Then in each small region we extract features such as color, texture, and shape, and all these characteristics are represented by a vector. So, an image can be represented as a three-dimension character array. On the basis of this three-dimension array, the local characteristics are conducted step by step to higher layer following the same mode of NR, and finally the FCNR can be obtained. Specific process is given below.

For any image $\mathbf{f} \subset \mathcal{I}_{Sq}$, we divide it into $M \times N$ rectangular blocks $\mathbf{f}_{ij}$ ($i = 1, 2, \ldots, M$; $j = 1, 2, \ldots, N$) with the same size using perpendicular lines network; that is,

$$\mathbf{f} = \begin{pmatrix} \mathbf{f}_{11} & \mathbf{f}_{12} & \cdots & \mathbf{f}_{1N} \\ \vdots & \vdots & \vdots & \vdots \\ \mathbf{f}_{M1} & \mathbf{f}_{M2} & \cdots & \mathbf{f}_{MN} \end{pmatrix}. \tag{1}$$

We extract some visual characteristics on each rectangular block in the same way. The details of features extraction methods will be presented in the third part of this paper. Normalizing the vectors with these characteristics as components and denoting them by $\mathbf{w}_{ij}$ ($i = 1, 2, \ldots, M$; $j = 1, 2, \ldots, N$), we can get an array

$$\mathbf{w}_{\mathbf{f}} = \begin{pmatrix} \mathbf{w}_{11} & \mathbf{w}_{12} & \cdots & \mathbf{w}_{1N} \\ \vdots & \vdots & \vdots & \vdots \\ \mathbf{w}_{M1} & \mathbf{w}_{M2} & \cdots & \mathbf{w}_{MN} \end{pmatrix} \tag{2}$$

which is the local feature representation of the image $\mathbf{f}$.

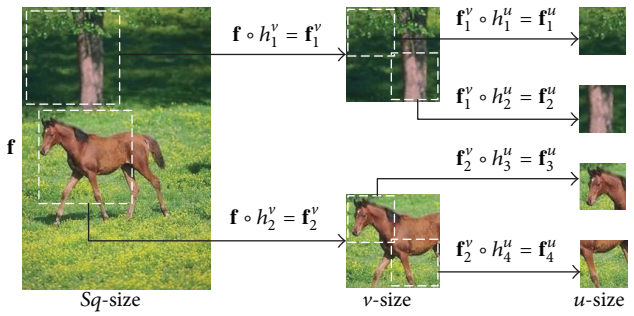

Figure 3: The hierarchical relationship of image and image patches.

It should be emphasized that these $\mathbf{w}_{ij}$ are all normalized vectors with the same dimension and each component of these vectors represents a feature of image block. An obvious advantage of normalization is said to be invariance to the brightness change of the image. If P characteristics are extracted from each rectangle image block $\mathbf{f}_{ij}$, then $\mathbf{w}_\mathbf{f}$ is a three-dimensional array and can be simply represented as

$$\mathbf{w}_\mathbf{f} = \left(w_{ijk}\right)_{M\times N\times P} \tag{3}$$

in which w_{ijk} $(i = 1,2,\ldots,M;\ j = 1,2,\ldots,N;\ k = 1,2,\ldots,P)$ denotes the kth feature of the image block in ith row and jth column of the image $\mathbf{f}$.

In this circumstance, the related notations and their meanings introduced in the previous section should be adjusted accordingly. The set of local feature representations of the images in the set $\mathcal{I}_{Sq}$ is denoted by $\mathcal{W}_{Sq^*}$; that is,

$$\mathcal{W}_{Sq^*} = \left\{\mathbf{w}_{\mathbf{f}_1}, \mathbf{w}_{\mathbf{f}_2}, \ldots, \mathbf{w}_{\mathbf{f}_{|\mathcal{I}_{Sq}|}}\right\}, \tag{4}$$

where Sq^* denotes the area of size $M \times N$. Accordingly, we use $\mathcal{W}_u = \{\mathbf{w}_1^u, \mathbf{w}_2^u, \ldots, \mathbf{w}_{|\mathcal{W}_u|}^u\}$ and $\mathcal{W}_v = \{\mathbf{w}_1^v, \mathbf{w}_2^v, \ldots, \mathbf{w}_{|\mathcal{W}_v|}^v\}$ to denote the sets of patches of size u and size v, respectively. It is needed to emphasize that the elements of previous $\mathcal{W}_u$ and $\mathcal{W}_v$ are obtained by sampling from the rows and columns of the array $\mathbf{w}_\mathbf{f}$ by moving windows rather than directly sampling from the image $\mathbf{f}$.

For example, assume that $\mathbf{f}$ is an image of size 256×384. It is divided into 8×8 square subblocks and we extract 6 characteristics from each subblock. At this point, the local feature representation $\mathbf{w}_\mathbf{f}$ is a three-dimensional array of size $32 \times 48 \times 6$ and the size of Sq^* is 32×48. If u-size is 15×15 and v-size is 21×21, then $\mathbf{w}^u \in \mathcal{W}_u$ and $\mathbf{w}^v \in \mathcal{W}_v$ are three-dimensional arrays of size $15 \times 15 \times 6$ and size $21 \times 21 \times 6$, respectively.

The notations $\mathcal{H}_u$ and $\mathcal{H}_v$ still denote the transformation sets of the transformations from u to v and v to Sq^*. The template sets $\mathcal{T}_u \subset \mathcal{W}_u$ and $\mathcal{T}_v \subset \mathcal{W}_v$ are obtained from $\mathbf{w}_\mathbf{f}$ in a similar way by moving the window on v and Sq^*, respectively. The elements in these template sets denoted by $\mathbf{t}_i^u$ $(i = 1,2,\ldots,|\mathcal{T}_u|)$ and $\mathbf{t}_j^v$ $(j = 1,2,\ldots,|\mathcal{T}_v|)$ are also some three-dimensional arrays.

Now we can define the feature conduction neural response. Firstly, assume $\mathbf{w}^v \in \mathcal{W}_v$ and, for any $h^u \in \mathcal{H}_u$, we have $\mathbf{w}^v \circ h^u \in \mathcal{W}_u$ according to Axiom 1. Taking a template $\mathbf{t}^u \in \mathcal{T}_u$, we call

$$\mathbf{N}_v\left(\mathbf{w}^v\right)\left(\mathbf{t}^u\right) = \max_{h^u\in\mathcal{H}^u} \sum_{i,j,k} \left(\left(\mathbf{w}^v \circ h^u\right) \cdot \mathbf{t}^u\right)_{ijk} \tag{5}$$

neural responses of $\mathbf{w}^v$ to the template $\mathbf{t}^u$, where $(\mathbf{w}^v \circ h^u) \cdot \mathbf{t}^u$ denotes a three-dimensional array that is obtained by multiplying the corresponding elements of two three-dimensional arrays $(\mathbf{w}^v \circ h^u)$ and $\mathbf{t}^u$, and $((\mathbf{w}^v \circ h^u) \cdot \mathbf{t}^u)_{ijk}$ represents the element of $(\mathbf{w}^v \circ h^u) \cdot \mathbf{t}^u$ in ith row and jth column of pth page. When $\mathbf{t}^u$ take over the template set $\mathcal{T}_u$, we can get a $|\mathcal{T}_u|$ dimensional vector

$$\mathbf{N}_v\left(\mathbf{w}^v\right) = \left(\mathbf{N}_v\left(\mathbf{w}^v\right)\left(\mathbf{t}_1^u\right), \ldots, \mathbf{N}_v\left(\mathbf{w}^v\right)\left(\mathbf{t}_{|\mathcal{T}_u|}^u\right)\right) \tag{6}$$

which is called the first layer of neural response of $\mathbf{w}^v$ to the template set $\mathcal{T}_u$. After normalization, it is denoted as $\widehat{\mathbf{N}}_v(\mathbf{w}^v)$; that is,

$$\widehat{\mathbf{N}}_v\left(\mathbf{w}^v\right) = \frac{\mathbf{N}_v\left(\mathbf{w}^v\right)}{\sqrt{\left\langle \mathbf{N}_v\left(\mathbf{w}^v\right), \mathbf{N}_v\left(\mathbf{w}^v\right)\right\rangle}}, \tag{7}$$

where $\langle\cdot,\cdot\rangle$ is inner product of two vectors in the usual sense.

Next, set $\mathbf{w}_f \in \mathcal{W}_{Sq^*}$, and, according to Axiom 1, we know that $\mathbf{w}_\mathbf{f} \circ h^v \in \mathcal{W}_v$. For any template $\mathbf{t}^v \in \mathcal{T}_v$, we call

$$\mathbf{N}_{Sq^*}\left(\mathbf{w}_\mathbf{f}\right)\left(\mathbf{t}^v\right) = \max_{h^v\in\mathcal{H}^v} \left\langle \widehat{\mathbf{N}}_v\left(\mathbf{w}_\mathbf{f} \circ h^v\right), \widehat{\mathbf{N}}_v\left(\mathbf{t}^v\right)\right\rangle \tag{8}$$

neural responses of $\mathbf{w}_\mathbf{f}$ to the template $\mathbf{t}^v$. When $\mathbf{t}^v$ take over the template set $\mathcal{T}_v$, we can get a $|\mathcal{T}_v|$ dimensional vector

$$\mathbf{N}_{Sq^*}\left(\mathbf{w}_\mathbf{f}\right) = \left(\mathbf{N}_{Sq^*}\left(\mathbf{w}_\mathbf{f}\right)\left(\mathbf{t}_1^v\right), \ldots, \mathbf{N}_{Sq^*}\left(\mathbf{w}_\mathbf{f}\right)\left(\mathbf{t}_{|\mathcal{T}_v|}^v\right)\right) \tag{9}$$

which is called the second layer of neural response of $\mathbf{w}_\mathbf{f}$ to the template set $\mathcal{T}_v$.

Finally, for any image $\mathbf{f} \in \mathcal{I}_{Sq}$, we define $\mathbf{N}_{Sq^*}(\mathbf{w}_\mathbf{f})$ as features conduction neural response (FCNR) of the image $\mathbf{f} \in \mathcal{I}_{Sq}$ and it is denoted by $\mathbf{N}(\mathbf{f})$; that is,

$$\begin{aligned}\mathbf{N}\left(\mathbf{f}\right) &= \mathbf{N}_{Sq^*}\left(\mathbf{w}_\mathbf{f}\right)\\ &= \left(\mathbf{N}_{Sq^*}\left(\mathbf{w}_\mathbf{f}\right)\left(\mathbf{t}_1^v\right), \ldots, \mathbf{N}_{Sq^*}\left(\mathbf{w}_\mathbf{f}\right)\left(\mathbf{t}_{|\mathcal{T}_v|}^v\right)\right).\end{aligned} \tag{10}$$

We add some remarks.

(i) The FCNR of an image is a vector whose dimension is equal to the number of templates in the second layer and has nothing to do with the dimension of the image itself. Therefore, in the process of image processing, we can transform all the images into vectors with the same dimension, regardless of the idea that the sizes of the images are the same or not.

(ii) Due to the use of image low-level visual features in the underlying layer, FCNR model effectively overcomes the shortcomings of pixel to pixel exhaustion algorithm of the NR model. At the same time, the low-level visual features of image are conducted to

the upper layer by the interaction between the subblocks of the image and the templates in every layer of the constructed hierarchical structure and make the FCNR contain high-level semantic elements of the image and this is very important in the task of CBIR.

(iii) From the perspective of learning theory, the feature extraction method of the FCNR belongs to the category of unsupervised learning [2, 19, 22], and the hierarchical structure is introduced to do deep learning for the low-level visual features.

2.3. Computational Complexity Analysis. In image retrieval task, we often have the real-time requirements. As a result, the complexity of the algorithm is very important when constructing the feature representation for CBIR. Here we analyze the computational complexity of the proposed method in this paper.

Consider the case of the n layers hierarchical architecture as shown in Figure 1. We define a set of global transformations where the range is always the entire image domain Sq rather than the next larger patch recursively setting

$$H_m^g = \left\{h : v_m \to Sq \mid h = h' \circ h'', \text{ with } h' \in H_{m+1}^g,\ h'' \in H_m\right\}, \tag{11}$$

for any $1 \le m \le n-1$, where H_n^g contains only the identity $\{I : Sq \to Sq\}$. In the above formula, we denote the transformation set from patch of mth layer to the next larger patch by H_m.

We denote the template in the mth layer by T_m and ignoring the cost of normalization and of precomputing the neural responses of the templates, the number of required operations to export the NR is given by

$$\tau = \sum_{m=1}^{n-1} \left(\left|H_m^g\right| \left|T_m\right| \left|T_{m-1}\right| + \left|H_{m+1}^g\right| \left|H_m\right| \left|T_m\right|\right), \tag{12}$$

where we denote for notational convenience the cost of computing the initial kernel by $|T_0|$ [22].

Because the image is preprocessed, $|H_m^g|$ in (12) in the calculation of the FCNR will be less than that in the calculation of the NR. This will eventually lead to the fact that τ of FCNR is far less than that of NR. In order to illustrate this point intuitively, we give a specific example.

Suppose $\mathbf{f}$ is an original image of size 256×384. In the calculation of the NR of $\mathbf{f}$, we take the u-size as 112×112 pixels and the v-size as 172×172 pixels. On the other hand, the image $\mathbf{f}$ will be divided into the square subblock of size 16×16 and 14 features will be extracted from each block before the calculation of the FCNR of $\mathbf{f}$. In this case, we take the u-size as 7×7 and the v-size as 11×11 which correspond to the 112×112 pixels and the 172×172 pixels in the original image. We also assume that the number of templates selected in each layer of the two methods is equal. Note that $n = 3$ in this paper; we can calculate τ of NR and FCNR using (12), respectively. The values of the parameters in (12) and the results τ of NR and FCNR are listed in Table 1.

Table 1: FCNR and NR are compared in terms of computational complexity.

	$\lvert H_1^g\rvert$	$\lvert H_2^g\rvert$	$\lvert H_3^g\rvert$	$\lvert H_1\rvert$	$\lvert H_2\rvert$	$\lvert T_0\rvert$	$\lvert T_1\rvert$	$\lvert T_2\rvert$	τ
NR	39585	18318	1	3721	18318	19246	500	300	4.18×10^{11}
FCNR	180	84	1	25	84	686	500	300	7.54×10^{7}

From Table 1, we can see that the number of operations of FCNR is less than one five-thousandth of the number of operations of NR. This means that the computational complexity of NR is much higher than that of FCNR. In fact, for the image with high resolution, it is not practical to directly calculate the NR. Usually, we will do some simple preprocessing for image before calculating its NR. Therefore, the difference of computation is not so great in practice (see Section 4).

3. CBIR System Based on FCNR

For a given image library, we divided all the images in the library into rectangular blocks with appropriate size using mutually perpendicular lines. In each rectangular block, the low-level features are extracted in the same way and thus the local feature representation of an original image is obtained. The local features representations of all images in the library will constitute a local characteristic database. So, we can construct a hierarchical architecture for the local feature representation and the template sets of every layer of the architecture can be obtained using the local characteristic database. On this basis and using the algorithm as mentioned in Section 2.2 to compute FCNR for all images in the library, we can establish a FCNR library associated with original image library. If a proper similarity measure is defined on the feature space of FCNR, then the image retrieval can be carried out.

When the user enters a query image for the relevant images retrieval, first of all, the user calculates the FCNR of the query image according to the above-mentioned steps and then calculates the similarity between the query image and all images in the image database according to the defined similarity measure. Finally, the user sorts the image in the library in decreasing order according to the similarity, and some numbers of images which are arranged in the top are output to the user. The flow diagram of CBIR based on FCNR is shown in Figure 4.

3.1. Local Low-Level Feature Extraction. In this paper, some simple and robust methods are used to extract fourteen basic low-level features, including color feature, texture feature, and shape feature, for the image block.

Similar to some CBIR related literatures, we use the well-known *YCbCr* color space in the extraction of color features [1, 8]. In this color space, the luminance information is stored with a single component Y, and the color information is stored with two color difference components Cb and Cr. We calculate the mean and standard deviation of Cr, Cb, and Y for each subblock, among which the mean values are denoted as g_1, g_2, and g_3 and the standard deviations are denoted

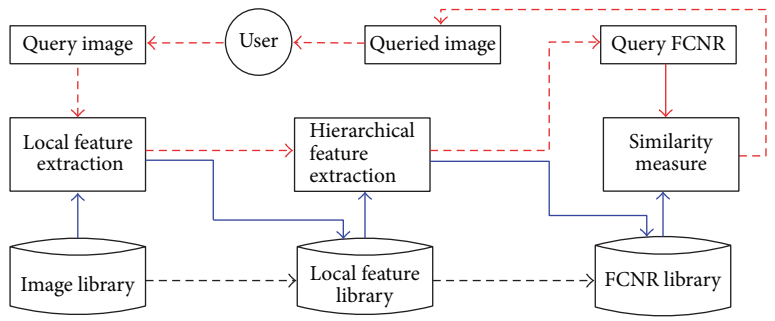

Figure 4: The flow diagram of CBIR based on FCNR.

as g_4, g_5, and g_6. In this way we can get six color features (for monochrome images, only two brightness features can be extracted).

Next, we will use Haar wavelet transform to extract texture features from the Y component of the rectangular image block. First of all, we will take Haar wavelet transform on each 4×4 subblock in the rectangular image block and four 2×2 matrixes can be obtained, which include a sampling approximation and three detail matrixes in three directions (horizontal, vertical, and diagonal). Set the three detail matrixes to be

$$\begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix}, \quad \begin{pmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{pmatrix}, \quad \begin{pmatrix} c_{11} & c_{12} \\ c_{21} & c_{22} \end{pmatrix}, \tag{13}$$

respectively, and let

$$a = \sqrt{\frac{\left(\sum_{i=1}^{2}\sum_{j=1}^{2} a_{ij}^2\right)}{4}}, \quad b = \sqrt{\frac{\left(\sum_{i=1}^{2}\sum_{j=1}^{2} b_{ij}^2\right)}{4}}, \quad c = \sqrt{\frac{\left(\sum_{i=1}^{2}\sum_{j=1}^{2} c_{ij}^2\right)}{4}}. \tag{14}$$

After wavelet transformation, we just assign the three variables to each pixel of the rectangular image block. Then, we can compute the averages and standard deviations of the three variables a, b, and c for each rectangular image block and denote the averages as g_7, g_8, and g_9 and the standard deviations as g_{10}, g_{11}, and g_{12}, respectively.

Note that the standard deviation g_4 of the Y component of the rectangular image block has been obtained; we take the thirteenth feature as

$$g_{13} = 1 - \frac{1}{\left(1 + (g_4)^2\right)}, \tag{15}$$

which is the smoothness of the image block and it reflects the relative smooth degree of brightness in the corresponding region. The last feature g_{14} is the entropy of the Y component of the rectangular image block; that is,

$$g_{14} = -\sum_{i=1}^{L-1} p(z_i) \log_2 p(z_i), \tag{16}$$

where $p(z)$ is the gray level histogram of the Y component of the rectangular image block and L is the number of possible gray series. Entropy is a measure of the randomness of the image elements [13].

In this way, the 14 features mentioned above are combined together; we can get low-level visual features representation of the rectangular image block, and we denote it as $\mathbf{g}$; that is,

$$\mathbf{g} = \{g_1, g_2, \ldots, g_{14}\}. \tag{17}$$

After obtaining the low-level visual features representation of all rectangle blocks of an image, we can get the local feature representation of the whole image as shown in (2).

3.2. The Similarity Measure. Retrieval accuracy is not only dependent on a robust feature representation, but also dependent on a good similarity measure. In order to highlight the advantages of FCNR in the image feature representation, we adopt a very basic and very natural way in this paper; that is, the similarity between two images is defined as the normalized inner product of their FCNRs. Specifically, for any of $\mathbf{f} \in \mathcal{I}_{Sq}$, its FCNR is a vector of $\mathbf{N}(\mathbf{f}) \in \mathbf{R}^{|\mathcal{T}_v|}$, where $|\mathcal{T}_v|$ represents the number of the templates in the second layer. To normalize $\mathbf{N}(\mathbf{f})$, we can obtain

$$\widehat{\mathbf{N}}(\mathbf{f}) = \frac{\mathbf{N}(\mathbf{f})}{\|\mathbf{N}(\mathbf{f})\|_{L^2(\mathbf{R}^{|\mathcal{T}_v|})}} \tag{18}$$

and the similarity of two images $\mathbf{f}, \mathbf{f}^* \in \mathcal{I}_{Sq}$ can be defined as

$$S(\mathbf{f}, \mathbf{f}^*) = \left\langle \widehat{\mathbf{N}}(\mathbf{f}), \widehat{\mathbf{N}}(\mathbf{f}^*) \right\rangle_{L^2(\mathbf{R}^{|\mathcal{T}_v|})}. \tag{19}$$

It is not difficult to see that the mode of definition of image similarity comes down in one continuous line of the definition of similarity of image patches at all layers in the process of construction of FCNR.

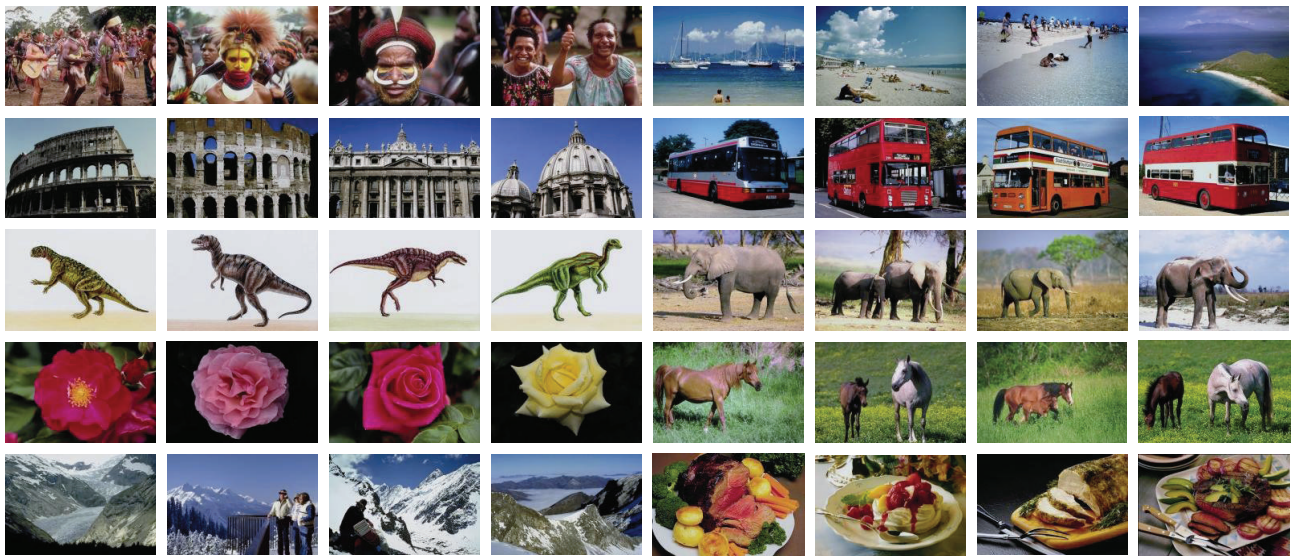

Figure 5: Example of images in COREL database.

Thus, when the user inputs the query image $\mathbf{q}$, the system firstly computes $\mathbf{N}(\mathbf{q})$ according to (10) and $\widehat{\mathbf{N}}(\mathbf{q})$ according to (18) and then calculates the similarities $S(\mathbf{q}, \mathbf{f}_i)$ ($i = 1, 2, \ldots, |\mathcal{I}_{Sq}|$) of the query image and all images in the database according to (18) and (19). Finally, descending sorts the images $\mathbf{f}_i$ based on the similarity $S(\mathbf{q}, \mathbf{f}_i)$ and outputs the top k images to the user as the query result, where the parameter k is specified by the user according to the query requirements.

4. Experiments

In this section, we will discuss simulation experiments to demonstrate the performance of the proposed method in image retrieval. Firstly, the evaluation standards of the performance of CBIR system are given and the appropriate parameters of FCNR method are selected. Then, we will compare the performance of the FCNR with the classical NR and the local neural responses (LNR) [17] in image retrieval. Finally, we also compare the proposed method with several feature extraction methods which were originally designed for image retrieval, including the benchmark method and some relatively new methods.

The image library used in the experiments contains 1000 images with size 256 × 384 or 384 × 256 selected from COREL database which is a general-purpose image database including about 60,000 pictures [1]. These selected images have ten classes, each of which has a semantic name and contains 100 pictures. For the sake of clarity, these 1000 images are numbered from 0 to 999. The semantic name and the corresponding number range of each class are listed in Table 2 [8]. We randomly selected four pictures from each class and show them in Figure 5.

It is necessary to emphasize that the templates used in the experiment for calculating FCNR are randomly intercepted from the local feature arrays of the image (and not from the original image) by moving or rotating window with specific size. The experiments were conducted on a computer with 4 GB random access memory and 2.60 GHz Intel(R) Core (TM) i5-3230 M processor, and the code was implemented in MATLAB in which the image processing toolbox functions are called [13].

4.1. Evaluation Standards and Parameter Determinations. There are a variety of ways for evaluation of the performance of retrieval. In this paper, we mainly use the recall-precision graph which is the most commonly used in community of image retrieval to evaluate the performance of FCNR. Precision P is defined as

$$P(k) = \frac{n_k}{k}, \tag{20}$$

where k is the number of retrieved images and n_k is the number of relevant images in the retrieved images. Recall R is defined as

$$R(k) = \frac{n_k}{N}, \tag{21}$$

where N is the number of all relevant images in the library. An optimal recall-precision graph would have a straight line; that is, precision is always at 1. Typically, when recall increases, precision decreases accordingly.

However, the results of one or two times of retrieval can not fully exhibit the advantage and disadvantage of an algorithm, and it is not convenient to compare with other methods. Therefore, we randomly selected 50 images from the image database to form a set of query images. For fixed recall, averaging the precision of the 50 time queries, we can obtain the recall-average precision graph which is a relatively reliable evaluation standard. In general, the high average precision and high recall mean that the algorithm has good performance. This means that the algorithm whose recall-average precision graph is over the right upper is better. In

Table 2: Ten classes of 1000 experimental images.

Classes	Semantic name	Number range
1	African	0~99
2	Beach	100~199
3	Building	200~299
4	Buses	300~399
5	Dinosaurs	400~499
6	Elephants	500~599
7	Flowers	600~699
8	Horses	700~799
9	Glaciers	800~899
10	Food	900~999

addition, due to the requirements of real-time in CBIR task, the shorter the query time the better the performance of the algorithm.

In our experiments, the images of size 384 × 256 are transformed into images of size 256 × 384 firstly through the rotation, and then all the images are divided into the square subblock size of 16 × 16, which is a total of 16 × 24 blocks for each image. For each image block, we extract local features by the methods described in Section 3.1 and we can get a three-dimensional array of 16 × 24 × 14 for every image. The templates sets are constructed by randomly extracting 500 patches of u-size and 300 patches of v-size, respectively, from the local feature arrays of some 10 images per class. In the process of constructing FCNR, two very important parameters are the size of u and v. In order to select the proper sizes, we have carried out a series of experiments for different sizes of u and v.

Figure 6 shows four recall-precision graphs corresponding to four different patch sizes. In these experiments, the number of retrieved images k is taken as 30. It is not difficult to see from Figure 6 that the sizes of u, and v are too small or too large to get good results. In contrast, when the u-size is 7×7 pixels and the v-size is 11×11 pixels, the retrieval results are the best. Therefore, we use these sizes in the experiments shown in Figure 6.

Figure 7 shows the top 20 images of two queries. The image in the front of the list is the query image and the figure at the bottom of each image is the number of the image in the library. As seen from Figure 7, the proposed CBIR system based on FCNR was done efficiently on the COREL image library. For the query semantics of "flower," the outputs of the top 20 images are all the theme of flowers, and these flowers take different color, size, background, and forms. This suggests that the high-level semantics of "flower" can be correctly identified by the system. It is worth mentioning that the two images of number 674 and number 677 are rotated before the local feature extraction and still can be retrieved, indicating that the FCNR algorithm preserves the rotation invariance of NR [24, 27, 28]. For the query semantics as "the elephant," the top 13 images of output are relevant to the query image, and among the top twenty images, only four images are inconsistent with the query image (the added border images in Figure 7).

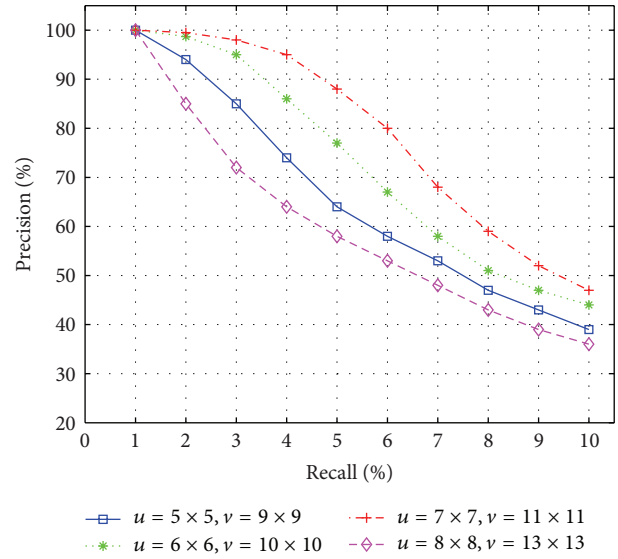

Figure 6: Recall-precision graphs of different patch sizes.

Table 3: Time consumption in three different methods.

Method	Local feature	Hierarchical feature	Query	Total
NR-based	0 s	1236.6 s	5.1 s	1241.7 s
LNR-based	0 s	1882.4 s	6.4 s	1888.8 s
FCNR-based	607.3 s	117.5 s	2.6 s	727.4 s

4.2. Comparison with NR-Based and LNR-Based Methods. Next, we will compare the performance of FCNR with the classic NR method and the LNR method in CBIR [17]. Local neural response is an improved version of the neural response, which uses sparse techniques in the representation of image and its subblocks. Before calculating NR and LNR, it is necessary to do preprocessing to the images as mentioned in Section 2.3. In order to be relatively fair, we adopt the approach which makes the algorithm perform best as reported in the relevant literature: we convert images into 60×90 gray images and the u-size is 15×15 pixels and the v-size is 21×21. We use a similar manner to select the template in the three methods, that is, to intercept 500 templates of u-size and 300 templates of v-size randomly from the gray images or the local feature matrix of images.

Table 3 shows the time consumption of the three different methods at different stages, and the retrieval performance is shown in Figure 8. In these experiments, the number of retrieved images k is still taken as 30.

It can be seen from Table 3 that both the hierarchical feature learning time and the query time of FCNR-based methods are significantly shorter than the other two methods. This is mainly because the latter two methods use exhaustion algorithm of translations of pixel to pixel. In particular, the LNR method, which introduces the solution of the quadratic optimization problems, is the most time-consuming method. Therefore, although the retrieval method based on FCNR can take some time on the extraction of local features, the time of

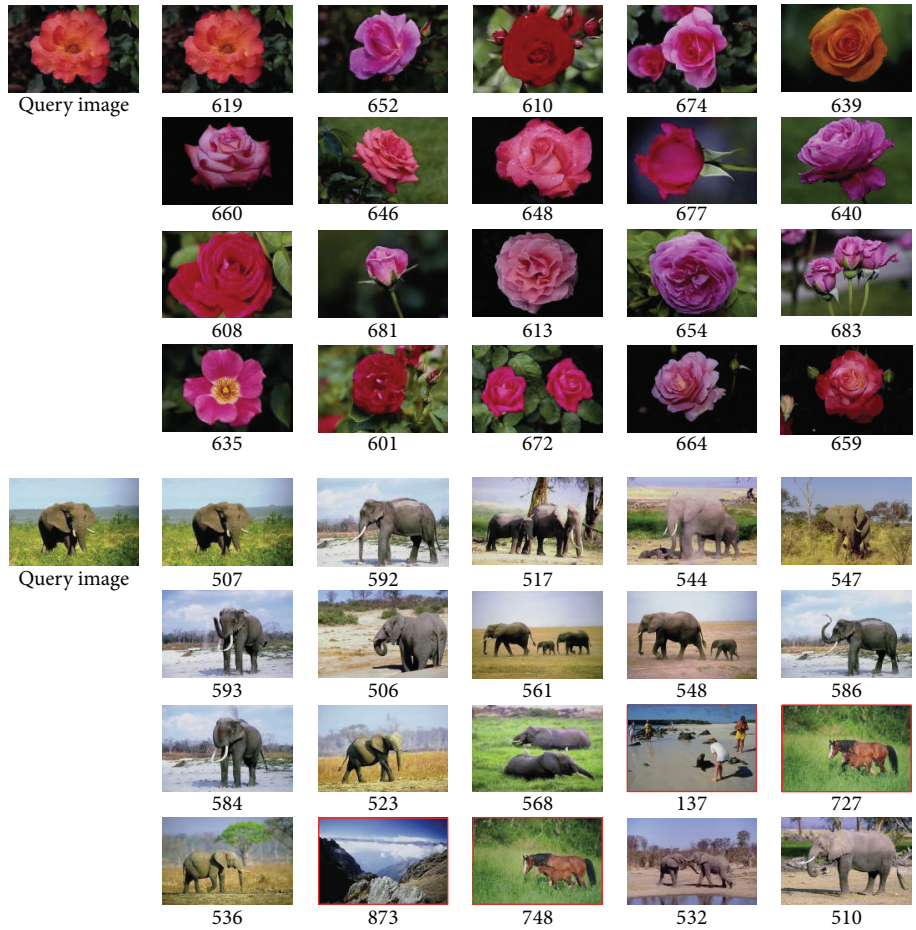

Figure 7: Two examples of FCNR-based query on COREL-1000 database.

learning the FCNR and the query time can be greatly reduced. This point is very important for the image retrieval task, because the real-time performance is the basic requirement in image retrieval [29].

Besides that, it is not hard to see from Figure 8 that the retrieval precision of the method based on FCNR is better than those methods based on NR or LNR. The main reason is that the FCNR-based method effectively overcomes the shortcomings of comparison of pixel values in the underlying image blocks just as in the NR and LNR methods. At the same time, the loss of color information also affected the performance of LNR and NR to a certain extent. By the way, the results based on LNR are better than that based on the NR. This is mainly because the localization and the sparse encoding in the LNR method make the image on the target location have a high value of neural response.

4.3. Comparison with Some Other Methods Proposed for Image Retrieval. Finally, we will compare FCNR with some other methods originally proposed for image retrieval, which include the edge histogram descriptor (EHD) method [10], the color difference histogram descriptor (CDH) [12], and the latest methods such as the error diffusion block truncation coding (EDBTC) [18] and the bandletized regions through support vector machines (BRSVM) [6].

As a benchmark, the EHD was initially used for texture image retrieval. In order to be fair, we extract features from the three components of R, G, and B of the image, and the edge intensity of the image block over 11 is used for the calculation of the histogram. Each component corresponds to an 80-dimensional feature vector, so that the resulting EHD feature vector is 240 dimensions. In the CDH representation we use *YCbCr* color space and the color and the orientation parameters are taken as 90 and 18, which is the relatively good performance parameter configuration as reported in the literature [12]. Thus the image of the CDH is expressed as a 108-dimension vector. The EDBTC produces two color quantizers and a bitmap image, which are further processed using vector quantization to generate the image feature descriptor. There are two features that are introduced in EDBTC, namely, color histogram feature and bit pattern histogram feature, to measure the similarity between a query image and the target image in database. The BRSVM method was prosed to overcome drawbacks and limitations of this traditional image segmentation technology. In BRSVM method, a bandelets transform based image representation technique is presented,

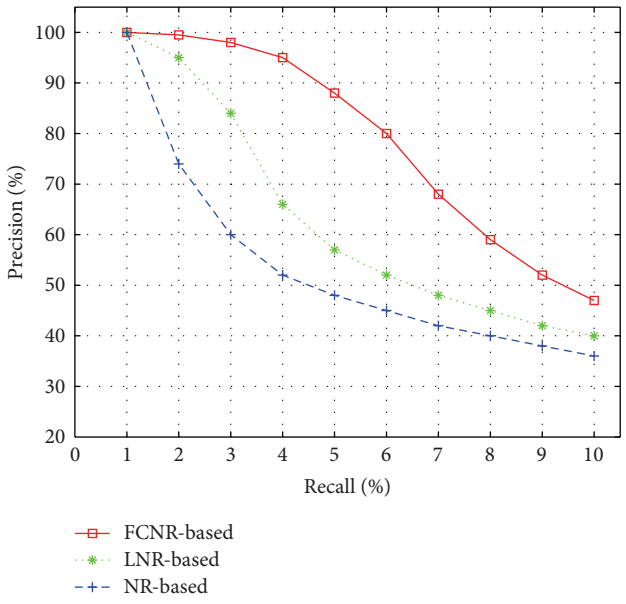

Figure 8: PR curves for three different algorithms.

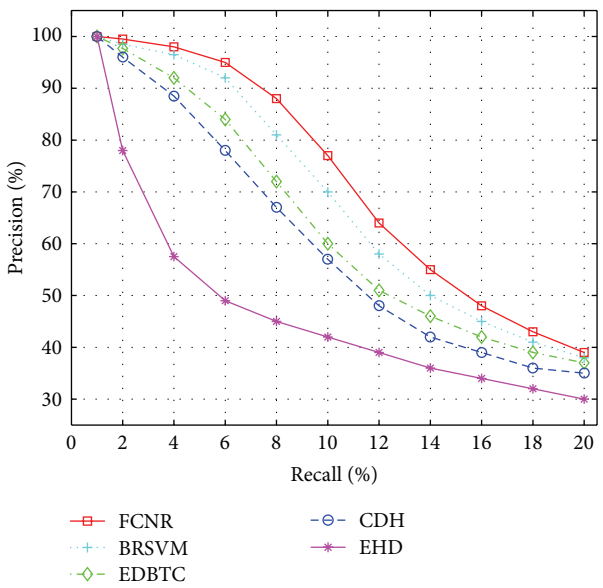

Figure 9: The PR curves of four different algorithms on COREL-1000 database.

which reliably returns the information about the major objects found in an image, and support vector machine is applied for image retrieval purposes.

Similarly, we randomly select 50 images from the image database to form a query image set Q, and the output image number is set to 10, 20, and 50, respectively. Table 4 lists the average precision and recall in the three different cases corresponding to the five methods, and Figure 9 shows the PR curve when the number of output images is set to 50.

It is not difficult to see from Table 4 and Figure 9 in the COREL-1000 image database that the performance of proposed method not only is significantly better than

Table 4: Average precision and recall of four different methods in different situations.

Output	Criterion	EHD	CDH	EDBTC	BRSVM	FCNR
10	Precision	48.47	62.44	63.87	67.13	68.63
	Recall	4.85	6.24	6.58	6.73	6.86
20	Precision	38.15	53.72	52.98	54.24	55.24
	Recall	7.63	10.74	10.64	10.87	11.05
50	Precision	29.76	42.87	44.23	45.79	46.98
	Recall	14.89	21.44	21.98	22.15	23.50

the MPEG-7 standard feature extraction method such as EHD but also has a stroke above those latest methods, such as EDBTC and BRSVM. From our point of view, this is mainly due to two reasons: the first reason is that, based on the hierarchical structure, FCNR is the result of deep learning on the low-level features of the image and the second reason is that the high-level semantic elements of images are integrated into the feature representation of FCNR by using the templates sets.

5. Conclusions

This research has been devoted to construct a new image feature representation of FCNR to be used for CBIR. It preserves the excellent characteristics of NR and LNR, such as being invariant to translation and illumination and robust to local distortion and clutter. More importantly, it also takes both visual feature and semantic feature into account in image recognition. The experimental results on the COREL-1000 image database have shown that, compared with NR and LNR, FCNR is more suitable for image retrieval tasks. In addition, the proposed method achieves a higher retrieval accuracy compared with other methods originally proposed for image retrieval in the COREL database. We attribute the effectiveness of the proposed method to both the local feature extraction and the hierarchical architecture which is used in deep learning of low-level visual features and take the high-level semantics of the image into its feature representation.

Although both theoretical analysis and experimental results show that FCNR is an applicable representation of image for CBIR, there are still some problems to be further studied.

(i) The template sets and the number of selected templates in this paper are determined according to experience; we do not pay any attention to qualitative and quantitative analysis. How to select the more representative templates and what are the optimal numbers of templates which are needed in the construction of FCNR are still important issues for future work [30].

(ii) In this paper, we used the simple inner product kernels as the similarity measure. In order to obtain a better retrieval performance, how to combine FCNR and the mainstream of similarity learning technology in recent years is worth studying problem.

(iii) As is well known, the relevance feedback technique plays an important role in image retrieval [9], and how to introduce relevance feedback into the algorithm proposed in this paper is another interesting subject.

Competing Interests

The authors declare that they have no competing interests.

Acknowledgments

This work is supported by Educational Commission of Hubei Province of China (D20131803), the Doctoral Scientific Research Fund of Hubei Automotive Industries Institute (BK201209), and the Youth Foundation of Hubei Automotive Industrial Institute (X2012XQ09).

References

[1] J. Z. Wang, J. Li, and G. Wiederhold, "SIMPLIcity: semantics-sensitive integrated matching for picture libraries," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 23, no. 9, pp. 947–963, 2001.

[2] V. N. Vapnik, *Statistical Learning Theory*, Adaptive and Learning Systems for Signal Processing, Communications, and Control, John Wiley & Sons, New York, NY, USA, 1998.

[3] T. Kohonen, *Self-Organizing Maps*, Springer, 1997.

[4] S. Tong and E. Chang, "Support vector machine active learning for image retrieval," in *Proceedings of the 9th ACM International Conference on Multimedia (MULTIMEDIA '01)*, pp. 107–118, ACM, Ottawa, Canada, 2001.

[5] L. Zhang, F. Lin, and B. Zhang, "Support vector learning for image retrieval," in *Proceedings of the International Conference on Image Processing*, vol. 2, pp. 721–724, Thessaloniki, Greece, October 2001.

[6] R. Ashraf, K. B. Bajwa, and T. Mahmood, "Content-based image retrieval by exploring bandletized regions through support vector machines," *Journal of Information Science and Engineering*, vol. 32, no. 2, pp. 245–269, 2016.

[7] P. Koikkalainen, "Fast deterministic self-organizing maps," in *Proceedings of the International Conference on Artificial Neural Networks (ICANN '95)*, Paris, France, 1995.

[8] S. Wu, M. K. M. Rahman, and T. W. S. Chow, "Content-based image retrieval using growing hierarchical self-organizing quadtree map," *Pattern Recognition*, vol. 38, no. 5, pp. 707–722, 2005.

[9] S. C. H. Hoi, M. R. Lyu, and R. Jin, "A unified log-based relevance feedback scheme for image retrieval," *IEEE Transactions on Knowledge and Data Engineering*, vol. 18, no. 4, pp. 509–524, 2006.

[10] B. S. Manjunath, J.-R. Ohm, V. V. Vasudevan, and A. Yamada, "Color and texture descriptors," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 11, no. 6, pp. 703–715, 2001.

[11] G.-H. Liu, L. Zhang, Y.-K. Hou, Z.-Y. Li, and J.-Y. Yang, "Image retrieval based on multi-texton histogram," *Pattern Recognition*, vol. 43, no. 7, pp. 2380–2389, 2010.

[12] G.-H. Liu and J.-Y. Yang, "Content-based image retrieval using color difference histogram," *Pattern Recognition*, vol. 46, no. 1, pp. 188–198, 2013.

[13] R. C. Gonzalez, R. E. Woods, and S. L. Eddins, *Digital Image Processing Using MATLAB*, Pearson Education India, 2004.

[14] K. Mikolajczyk and C. Schmid, "A performance evaluation of local descriptors," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 27, no. 10, pp. 1615–1630, 2005.

[15] Y. Ke and R. Sukthankar, "PCA-SIFT: a more distinctive representation for local image descriptors," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pp. 506–513, June-July 2004.

[16] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *International Journal of Computer Vision*, vol. 60, no. 2, pp. 91–110, 2004.

[17] H. Li, Y. Wei, L. Li, and C. L. P. Chen, "Hierarchical feature extraction with local neural response for image recognition," *IEEE Transactions on Cybernetics*, vol. 43, no. 2, pp. 412–424, 2013.

[18] J.-M. Guo, H. Prasetyo, and J.-H. Chen, "Content-based image retrieval using error diffusion block truncation coding features," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 25, no. 3, pp. 466–481, 2015.

[19] T. Serre, L. Wolf, S. Bileschi, M. Riesenhuber, and T. Poggio, "Robust object recognition with cortex-like mechanisms," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 29, no. 3, pp. 411–426, 2007.

[20] M. Ursino and G. E. La Cara, "A model of contextual interactions and contour detection in primary visual cortex," *Neural Networks*, vol. 17, no. 5-6, pp. 719–735, 2004.

[21] K. Huang, D. Tao, Y. Yuan, X. Li, and T. Tan, "Biologically inspired features for scene classification in video surveillance," *IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics*, vol. 41, no. 1, pp. 307–313, 2011.

[22] S. Smale, L. Rosasco, J. Bouvrie, A. Caponnetto, and T. Poggio, "Mathematics of the neural response," *Foundations of Computational Mathematics*, vol. 10, no. 1, pp. 67–91, 2010.

[23] H. Li, Y. Wei, L. Li, and Y. Yuan, "Similarity learning for object recognition based on derived kernel," *Neurocomputing*, vol. 83, pp. 110–120, 2012.

[24] Y. Y. Tang, T. Xia, Y. Wei, H. Li, and L. Li, "Hierarchical kernel-based rotation and scale invariant similarity," *Pattern Recognition*, vol. 47, no. 4, pp. 1674–1688, 2014.

[25] Z. Hu and H. Xiao, "Soft sparse coding neural response for image feature extraction," *Optik—International Journal for Light and Electron Optics*, vol. 126, no. 17, pp. 1510–1519, 2015.

[26] T. Serre, M. Kouh, C. Cadieu, U. Knoblich, G. Kreiman, and T. Poggio, "A theory of object recognition: computations and circuits in the feed forward path of the ventral stream in primate visual cortex," Tech. Rep., DTIC Document, 2005.

[27] S. Smale, T. Poggio, A. Caponnetto, and J. Bouvrie, *Derived Distance: Towards a Mathematical Theory of Visual Cortex*, 2007.

[28] H. Li, Y. Wei, L. Li, and Y. Y. Tang, "Infrared moving target detection and tracking based on tensor locality preserving projection," *Infrared Physics & Technology*, vol. 53, no. 2, pp. 77–83, 2010.

[29] Y. Rui, T. S. Huang, and S.-F. Chang, "Image retrieval: current techniques, promising directions, and open issues," *Journal of*

Visual Communication and Image Representation, vol. 10, no. 1, pp. 39–62, 1999.

[30] Z. Hu, H. Xiao, and X. Zhen, "Template selection algorithm for NR-based image feature extraction," in *Proceedings of the 10th International Conference on Natural Computation (ICNC '14)*, pp. 1143–1147, IEEE, August 2014.