Multidim Syst Sign Process (2017) 28:1071–1089
DOI 10.1007/s11045-016-0386-3

Robust discriminative extreme learning machine for relevance feedback in image retrieval

Shenglan Liu[1,2] · **Lin Feng**[1,2] · **Yang Liu**[1,2] · **Jun Wu**[1,2] · **MuXin Sun**[1,2] · **Wei Wang**[1,2]

Received: 16 September 2015 / Revised: 1 February 2016 / Accepted: 11 February 2016 / Published online: 26 February 2016

Abstract Relevance feedback (RF) has long been an important approach for multi-media retrieval because of the semantic gap in image content, where SVM based methods are widely applied to RF of content-based image retrieval. However, RF based on SVM still has some limitations: (1) the high dimension of image features always make the RF time-consuming; (2) the model of SVM is not discriminative, because labels of image features are not sufficiently exploited. To solve above problems, we proposed robust discriminative extreme learning machine (RDELM) in this paper. RDELM involved both robust within-class and between-class scatter matrices to enhance the discrimination capacity of ELM for RF. Furthermore, an angle criterion dimensionality reduction method is utilized to extract the discriminative information for RDELM. Experimental results on four benchmark datasets (Corel-1K, Corel-5K, Corel-10K and MSRC) illustrate that our proposed RF method in this paper achieves better performance than several state-of-the-art methods.

Keywords Relevance feedback · Image retrieval · Extreme learning machine · Robust discriminative information

This work was supported by National Natural Science Foundation of P. R. China (61173163, 61370200) and China Postdoctoral Science Foundation (ZX20150629).

✉ Lin Feng
fenglin@dlut.edu.cn

Shenglan Liu
liusl@mail.dlut.edu.cn

1 Faculty of Electronic Information and Electrical Engineering, Dalian University of Technology, Dalian 116024, China

2 School of Innovation Experiment, Dalian University of Technology, Dalian 116024, China

1 Introduction

Image retrieval is an effective approach to search the special images for queries, and has many emerging applications. There are a significant number of methods for extracting image features which play an important role in CBIR (Swain and Ballard 1991; Liu et al. 2011; Ojala et al. 2002). Image representation methods mainly fall into three categories: color, texture and shape feature. Liu et al. (2011) propose Micro-structure Descriptor (MSD) which used edge orientation to extract micro-structures of images. It depends on the similarity of underlying colors and structure element correlation, then a 72-length feature vector can be obtained for describing an image. Ojala et al. (2002) propose local binary patterns (LBP) as texture descriptor of an image, which is widely applied to face recognition (Tan and Triggs 2007), image retrieval (Murala and Wu 2014), etc.

The extracting methods of image feature have improved the image retrieval results, but do not always describe the semantic of images. Consequently, RF becomes an essential technique to deal with the semantic gap (see Fig. 1) (Feng et al. 2015). Yong et al. first involved the RF concept into a CBIR system. Then, many methods of RF have been proposed to enhance image retrieval results achieved by Non-RF based methods while semantic gap existed. He (2004) proposed a semi-supervised learning approach to keep the local space information of the query image by calculating the local data basis of the query image from RF. Moreover, by utilizing Locality Preserving Projections (LPP) algorithm (He and Niyogi 2003), the local visual features were extracted and the graph of LPP were reconstructed with RF. Kundu et al. (2015) proposed graph-based relevance feedback mechanism for image retrieval task. Zhang et al. (2015) proposed query special rank fusion to solve the ranking problem in image retrieval which could also help RF to improve the retrieval results. Besides, many SVM active learning (Hoi et al. 2008; Anitha and Rinesh 2013; Lu et al. 2006; Hoi and Lyu 2005; Wang et al. 2011) algorithms have been applied in RF-CBIR. However, SVM based algorithms are difficult to obtain exciting classification results, for the labeled sample size is generally small in RF-CBIR. Hoi et al. (2008) proposed a semi-supervised SVM batch mode active learning algorithm. The supervised SVM classifier is trained by constructing a kernel function learned from the Laplacian graph which represents the global geometry structure of the samples. And then the supervised SVM classifier is used to label samples containing most information. By using the integrated SVM learner in RF-CBIR, Wang et al. (2011) handled the unbalance problem between positive and negative samples with the asymmetric bagging SVM and avoided the outfitting problem with subspace random chosen.

Fig. 1 Semantic gap in image retrieval

ELM is an efficient classification method. There are many ELM-based modified versions in recent years (Huang et al. 2006, 2012; Huang and Chen 2007; Horata et al. 2012; Deng et al. 2009; Liu et al. 2014; Tang and Han 2009; Zhang and Yang 2015; Huang 2015; Cao and Lin 2015; Cao et al. 2015a, b, 2012). Huang et al. proved that by adding activation functions as nodes randomly, SLFNs can approximate any continuous target function, such as, RBF. On this basis, they proposed incremental ELM algorithm (Huang et al. 2006) and convex incremental ELM algorithm (Huang and Chen 2007). Outliers of the training datasets have a great impact on the precision of ELM (Zhang and Luo 2015). Considering this, Horata et al. (2012) proposed the Robust ELM algorithm, which gets the output weights by constructing an Extended Complete Orthogonal Decomposition. The final output weights are computed by iterating the initial output weights. Besides, the time complexity of Robust ELM is equal to that output weights computed by Singular Value Decomposition (SVD). Its main reason is that the empirical risk of ELM is lowered by the Robust ELM algorithm through the reasonable numerical computation. Hence, we can conclude that the robustness of ELM is closely related to the empirical risk. Deng et al. (2009) proposed a weighted Regular ELM (RELM) algorithm. This algorithm enhances the robustness of RELM to outliers and by lowering the empirical risk weights of outliers. Even if the methods introduced above have done a lot of improvements on ELM, but there are always few labeled samples in real-world datasets. In order to take full advantage of unlabeled samples and improve the generalization capability of ELM, Liu et al. (2014) proposed a semi-supervised ELM, which extends the ELM algorithm to a semi-supervised version based on the graph theory. In addition to the above theoretical improvements on ELM, some applications of ELM-related methods also have been proposed. For instance, Iosifidis et al. proposed (Tang and Han 2009) minimum class variance extreme learning machine method for human action recognition [also see related papers (Iosifidis et al. 2013, 2014; Mohammed et al. 2011)]. Akusok et al. utilized optimally pruned ELM to analysis web content (Minhas et al. 2010). Jin et al. (2015) proposed an Ensemble based extreme learning machine method for face matching.

There are few image retrieval researches related to ELM and discriminative methods. From previous works, we know that the data distribution in ELM feature space plays an important role on clustering (Akusok et al. 2015) and classification (Tang and Han 2009). Motivated by this idea, in this paper, we exploit a novel ELM-based method called RDELM for RF in CBIR system (see Fig. 2). The main contributions of RDELM are as follows:

(1) RDELM utilizes within and between scatters in hidden layer to detect a discriminative classification model.
(2) The model of RDELM can balance the within and between scatters of feature space by using cosine metric.
(3) RDELM involves supervised reduction with angle criterion in ELM feature space, which makes the proposed classification method stable and robust in image retrieval task.

The rest of the paper is organized as follows. ELM classification method is reviewed in Sect. 2. Section 3 analyzes the scatters in ELM feature space, and proposes a supervised dimensionality reduction method. In Sect. 4, we introduce the RDELM method for RF in CBIR system. Experimental results are shown in Sect. 5. The last section is the conclusion of this paper.

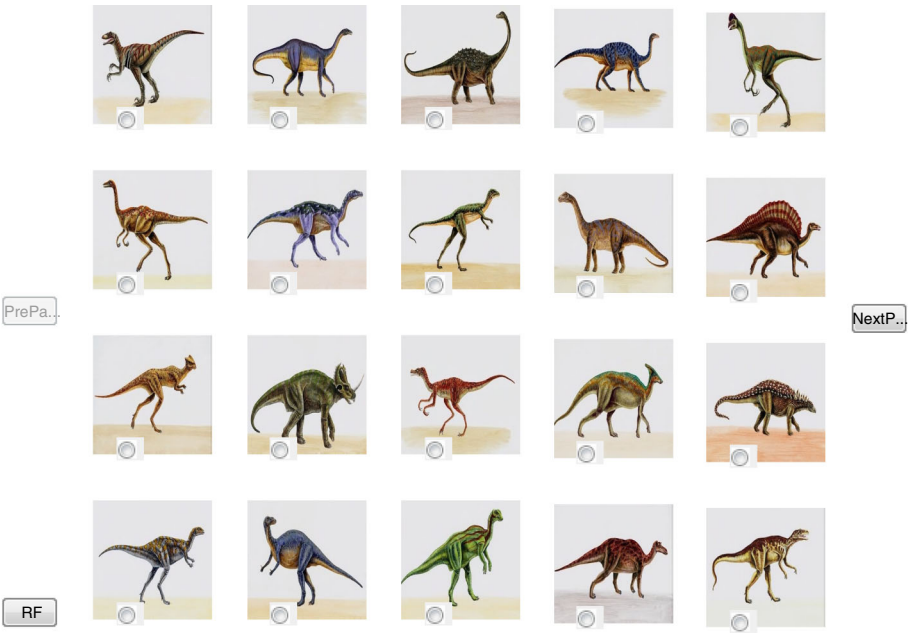

Fig. 2 RDELM for relevance feedback in CBIR system

2 Brief of ELM

Given D dimensional dataset $X = [x_1, x_2, \ldots, x_N]^T \in R^{D\times N}$ with N samples , and corresponding labels $t_i = [t_{i1}, t_{t2}, \ldots, t_{im}]^T \in R^m$, with $\tilde{N}$ hidden neurons in network, then the output of the network can be obtained by the following equation

$$\sum_{i=1}^{\tilde{N}} \beta_i g(a_i x_j + b_j) = o_j \tag{1}$$

where $j = 1, \ldots, N$, $g(a_i x_j + b_j)$ denotes activation function in Neural Networks (NN), $a_i = [a_{i1}, a_{i2}, \ldots, a_{iN}]^T$ is the weight vector connecting input neurons and the ith hidden node, b_i is bias of the ith hidden node.

For hidden neurons $\tilde{N}$, Eq. (1) can be rewritten as a matrix format: $H\beta = T$, where network hidden layer output matrix can be expressed as $H = \begin{bmatrix} g(a_1, x_1, b_1) & \cdots & g(a_{\tilde{N}}, x_1, b_{\tilde{N}}) \\ \vdots & \cdots & \vdots \\ g(a_1, x_N, b_1) & \cdots & g(a_{\tilde{N}}, x_N, b_{\tilde{N}}) \end{bmatrix}_{N\times\tilde{N}} = \begin{bmatrix} h_1 \\ \vdots \\ h_N \end{bmatrix}$, $\beta = \begin{bmatrix} \beta_1^T \\ \vdots \\ \beta_{\tilde{N}}^T \end{bmatrix}_{\tilde{N}\times m}$ and $T = \begin{bmatrix} t_1^T \\ \vdots \\ t_N^T \end{bmatrix}_{N\times m}$, $h_i = \left[g(a_1, x_i, b_1), \ldots, g(a_{\tilde{N}}, x_i, b_{\tilde{N}})\right]$, $i = 1, 2, \ldots, N$. The standard single hidden feedforward neural networks (SLFNs) is to compute appropriate $\hat{a}_i$, $\hat{b}_i$ and $\hat{\beta}(i = 1, \ldots, \tilde{N})$ to satisfy:

$$\left\| H\left(\hat{a}_1, \ldots, \hat{a}_{\tilde{N}}, \hat{b}_1, \ldots, \hat{b}_{\tilde{N}}\right)\beta - T \right\|_2^2 = \min_{a_i, b_i, \beta} \left\| H\left(a_1, \ldots, a_{\tilde{N}}, b_1, \ldots, b_{\tilde{N}}\right)\beta - T \right\|_2^2 \tag{2}$$

Equation (2) can be solved by gradient-based methods. Liu et al. (2011) have proved that the weights between input layer and the biases need no adjustment compared with the standard SLFNs. The input weight a_i and the biases b_i of hidden layer in ELM can be selected randomly, if the activation function of network is continuously differentiable. From Eq. (2), it is clear that the solution of SLFNs' network can be get by the least-squares methods of the linear system $H\beta = T$.

If the number of hidden neurons $\tilde{N}$ is equivalent to the number of training samples N, that is $\tilde{N} = N$, the matrix H is invertible directly. However, in most cases $\tilde{N} \neq N$, the solution of $H\beta = T$ can be written as

$$\hat{\beta} = H^{\dagger}T \tag{3}$$

where $H^{\dagger}$ denotes the generalized inverse matrix of H. $H^{\dagger}$ can be calculated by SVD or least-squares. Meanwhile, in order to avoid outfitting, we hope to keep the weights as small as possible. Thus, the optimization model of ELM can be modeled as follows:

$$\begin{aligned} &\min_{\beta} \frac{1}{2}\|\beta\|_2^2 + \frac{1}{2}C\sum_{i=1}^{N}\xi_i^2 \\ &s.t.\ \sum_{i=1}^{\tilde{N}}\beta_i g(a_i \cdot x_j + b_j) - t_j = \xi_j,\ \ j = 1, 2, \ldots, N \end{aligned} \tag{4}$$

where C is a penalty factor, which can balance the empirical risk and structural risk. The optimization model in Eq. (4) can be transformed to an unconstrained optimization problem of solving β with Lagrange method.

ELM can be simply implemented as follows:

(1) Assign arbitrary input weights a_i and bias b_i, $i = 1, \ldots, \tilde{N}$;
(2) Calculate the hidden layer output matrix H by active function;
(3) Compute the output weight $\hat{\beta} = H^{\dagger}T$, where $T = [t_1, \ldots, t_N]^T$.

3 Robust supervised dimensionality reduction for discriminative extreme learning machine

In this section, we give a theoretical analysis of the within and between scatter in ELM hidden layer matrix $H = [h_1, \ldots, h_N]^T$. Then, we introduce supervised dimensionality reduction in ELM feature space if $\tilde{N} < N$, otherwise $\tilde{N} \geq N$, a dimensionality reduction free method will be proposed in Sect. 4.

Supposed that matrix H is grouped as $H = [H_1, H_2, \ldots, H_m]^T$, where $H_i \in \mathrm{R}^{\tilde{N} \times N_i}$ is the data matrix consisting of data from the ith class with a sample size of N_i satisfying $N = \sum_{i=1}^{m} N_i$. The mean vector of the ith class can be defined as $\mu_h^i = \frac{1}{N_i}\sum_{j=1}^{N_i} h_j^i$, and the mean vector of H can be calculated as $\mu_h = \frac{1}{N}\sum_{i=1}^{N} h_i$.

In linear discriminant analysis (LDA), the minimization of within-class scatter corresponding subspace is $W_{opt}^{LDA} = \arg\min \sum_{i,j} \left\| WW^T\left(h_j^i - \mu_h^i\right)\right\|_2^2$, where $WW^T(\cdot)$ is the orthogonal projection of $(\cdot)$. $(\cdot)_e$ denotes the unit vector of $(\cdot)$. Then we can compute the cosine value of $\alpha_j^i = \left\langle WW^T\left(h_j^i - \mu_i^h\right), h_j^i - \mu_i^h\right\rangle$ as follows:

$$\cos^2\alpha_j^i = \frac{\left[\left(h_j^i-\mu_h^i\right)^T\cdot WW^T\left(h_j^i-\mu_h^i\right)\right]^2}{\left\|h_j^i-\mu_h^i\right\|_2^2\left\|WW^T\left(h_j^i-\mu_h^i\right)\right\|_2^2} = \left(h_j^i-\mu_h^i\right)_e^T\cdot WW^T\left(h_j^i-\mu_h^i\right)_e \tag{5}$$

From formula (5) we can see that minimization of within-class scatter equals to $\min_W \sum_{i,j}\cos^2\alpha_j^i$ in geometry. Similarly, the maximization of between-class scatter equals to $\max_W \sum_{i,j}\cos^2\beta_h^i$ in geometry, where $\beta_h^i = \left\langle WW^T\left(h_i-\mu_h\right), h_i-\mu_h\right\rangle$.

$$\sum_{i=1}^m\left[\frac{N_i}{N}\frac{1}{N_i}\sum_{j=1}^{N_i}\cos^2\alpha_j^i\right] = tr\left[\frac{1}{N}W^T\left[\sum_{i=1}^m\sum_{j=1}^{N_i}\left(h_j^i-\mu_h^i\right)_e^T\left(h_j^i-\mu_h^i\right)_e\right]W\right] = tr\left[W^TS_w'W\right] \tag{6}$$

where $S_w' = \frac{1}{N}\sum_{i=1}^m\sum_{j=1}^{N_i}\left(h_j^i-\mu_h^i\right)_e^T\left(h_j^i-\mu_h^i\right)_e = H_{we}^TH_{we}$,
$H_{we} = \frac{1}{\sqrt{N}}\left[\left(h_1^1-\mu_h^1\right)_e,\ldots,\left(h_1^{N_1}-\mu_h^1\right)_e,\ldots,\left(h_m^1-\mu_h^m\right)_e,\ldots,\left(h_m^{N_m}-\mu_h^m\right)_e\right]^T$.

The between-class scatter can be written as:

$$\sum_{i=1}^m\left(\frac{N_i}{N}\cos^2\beta_h^i\right) = \frac{1}{N}\sum_{i=1}^m\frac{N_i\left(\mu_h^i-\mu_h\right)WW^T\left(\mu_h^i-\mu_h\right)^T}{\left\|\mu_h^i-\mu_h\right\|_2^2} = tr\left[\frac{1}{N}W^T\sum_{i=1}^m N_i\left(\mu_h^i-\mu_h\right)_e^T\left(\mu_h^i-\mu_h\right)_eW\right] = tr\left[W^TS_b'W\right] \tag{7}$$

where $S_b' = \frac{1}{N}\sum_{i=1}^m N_i\left(\mu_h^i-\mu_h\right)_e^T\left(\mu_h^i-\mu_h\right)_e = H_{be}^TH_{be}$,
$H_{be} = \frac{1}{\sqrt{N}}\left[\sqrt{N_1}\left(\mu_h^1-\mu_h\right)_e,\ldots,\sqrt{N_m}\left(\mu_h^m-\mu_h\right)_e\right]^T$.

Then, the supervised dimensionality reduction function based on angle criterion can be expressed as follows:

$$F(W) = \arg\min_w\ tr\left[W^T\left(S_w'-S_b'\right)W\right] \tag{8}$$

The constraint of Eq. (8) is $W^TW = I$, where I is an identity matrix.

S_w' and S_b' can balance the scatter of within and between class. This conclusion is shown in Fig. 3. There are three classes features in Fig. 3a, b. In Fig. 3a, LDA can obtain an appropriate subspace, but fails in Fig. 3b. Comparing with LDA, our angle criterion approach is suitable in the two feature spaces because of the scatter balance. The conclusion is detailed in reference Liu et al. (2015).

4 Robust discriminative extreme learning machine for RF

In this section, we will propose the RDELM model for classification, and give a RDELM-based relevance feedback method for CBIR system. RDELM is introduced in Sect. 4.1,

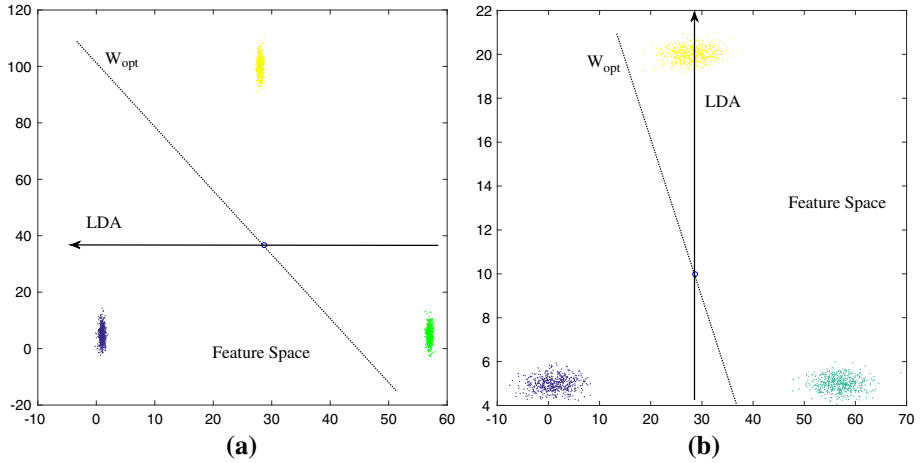

Fig. 3 Comparing our robust scatter method to LDA in ELM hyperplane selection. **a** Feature space data1, **b** feature space data2

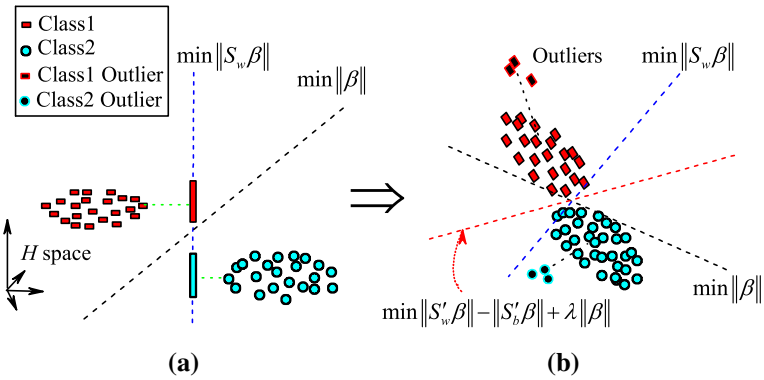

Fig. 4 Comparison of hyper plane on different versions ELM. **a** MCVELM and ELM, **b** RDELM, MCVELM and ELM

and both dimensionality reduction based RDELM and non-dimensionality reduction based RDELM methods are included. In Sect. 4.2, the RDELM-based retrieval method is utilized on relevance feedback technique for CBIR system.

4.1 Robust discriminative extreme learning machine

We know that the original ELM utilizes H to compute the output weights β directly. Actually, there are some important structures in hidden layer feature space of ELM. For example, He et al. (2014) proposed clustering method in ELM feature space. When training samples containing outliers, over-fitting problem is important to the classifier (see Fig. 4b).

As shown in Fig. 4, original ELM and MCVELM can get appropriate decision spaces when there is no outlier involved in feature space (Fig. 4a). But if outliers exist because of over-fitting in the feature space, the original ELM and MCVELM may generate unsuitable

decision spaces (Fig. 4b). Consequently, we can involve the robust class structure to construct an ELM-based classifier.

4.1.1 Dimensionality reduction free based version of RDELM

Considering the minimization of within-class scatter, the maximization of between-class scatter and the algorithm robustness, RDELM method can be modeled by solving an optimization problem as follows:

$$\begin{aligned} &\min_{\beta} \|H_{we}\beta\|_2^2 - \|H_{be}\beta\|_2^2 + \lambda\|\beta\|_2^2 + C\sum_{i=1}^{N}\|\xi_i\|_2^2 \\ &s.t. \quad h_i\beta = t_i - \xi_i, \quad i = 1, 2, \ldots, N \end{aligned} \tag{9}$$

In (9), $\|H_{we}\beta\|_2^2$ and $\|H_{be}\beta\|_2^2$ aim to balance the class structure of the feature space in ELM, $\lambda\|\beta\|_2^2$ is to avoid over-fitting problem, and the $C\sum_{i=1}^{N}\|\xi_i\|_2^2$ term can minimize the error which is generated by the constraint. RDELM is robust which is analyzed in Sect. 3. The solution of (9) can be calculated by Lagrange multiplier method. β can be expressed by

$$\beta = \left(\frac{1}{C}\left(S'_w - S'_b\right) + \frac{1}{C\lambda}I + H^T H\right)^{-1} H^T T \tag{10}$$

From (10), the output function of RDELM classifier can be written as:

$$f(x) = h(x)\beta = h(x)\left(\frac{1}{C}\left(S'_w - S'_b\right) + \frac{1}{C\lambda}I + H^T H\right)^{-1} H^T T \tag{11}$$

4.1.2 Dimensionality reduction based version of RDELM

When $\tilde{N} < N$, $P = \frac{1}{C}\left(S'_w - S'_b\right) + \frac{1}{C\lambda}I + H^T H$ may not be invertible. To keep the discriminative information and make P invertible, the dimensionality reduction method in Eq. (8) can be applied to the basis RDELM classifier. Then, the within-class and between scatter matrices are rewritten by $\hat{S}_w = W^T S'_w W$ and $\hat{S}_b = W^T S'_b W$, respectively. So, the final solution of dimensionality reduction based version of RDELM can be got by

$$\beta = \left[W^T\left(\frac{1}{C}\left(S'_w - S'_b\right) + \frac{1}{C\lambda}I + H^T H\right)W\right]^{-1} W^T H^T T \tag{12}$$

where W has been calculated by Eq. (8) in Sect. 3. We illustrate that $\tilde{N}$ is chosen by users. Then, we will know whether $\tilde{N} < N$ or not, and select the method in Sects. 4.1.1 or 4.1.2.

4.2 The RF algorithm using RDELM

The scheme of relevance feedback based on SVM has been exploited in CBIR system. However, an unsatisfied performance may be occurred because of the over-fitting problem when the large number of feature dimensions need to be calculated in SVM. Furthermore, SVM has to generate a lot of support vectors which leads to classification a time-consuming work. In order to solve the above problems, a RDELM-based relevance feedback method is proposed in CBIR system (see Fig. 5).

According to the above introduction, the retrieved images in the previous step can be marked by users themselves. Take the retrieved images as training data, a two classes learning algorithm can be carried out by ELM, based on which a classifier can be constructed. Given

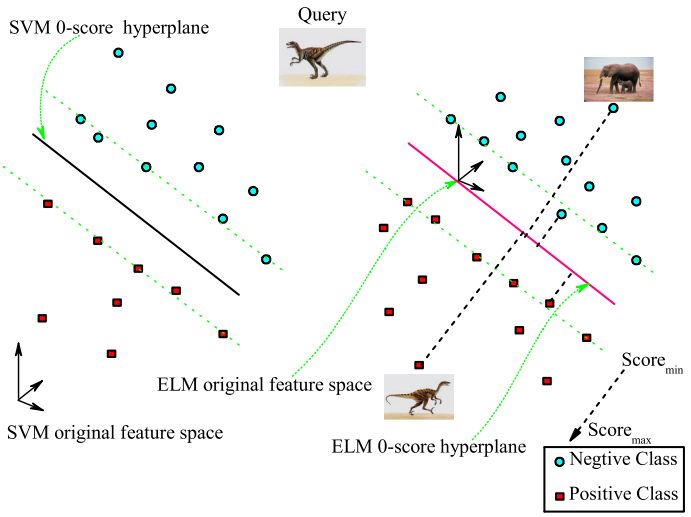

Fig. 5 Comparison of ELM-based and SVM-based RF image retrieval

an arbitrarily distinct sample set (x_i, t_i), $x_i \in \mathrm{R}^D$, where x_i represents feature vector of an image containing the information of color, texture and shape, and t_i is the label of x_i, $t_i \in \{(1,-1)^T, (-1,1)^T\}$. Then constructing a classifier by ELM algorithm, and classifying rest images into relevant class or irrelevant class. On this basis, we can obtain a better result after several feedbacks after sorting images according to their relevance to the query. The process of our algorithm is summarized in Algorithm 1:

Algorithm 1: RF based on RDELM	
Step 1.	Obtain top N images of the total images by a traditional retrieval algorithm;
Step 2.	Distinct the N images into two classes, and mark them as relevant set I^+ and irrelevant set;
Step 3.	Prepare the training data $(x_i, t_i)$for the RDELM algorithm, where $x_i \in I^+ \cup I^-$ and $t_i \begin{cases} (1,-1)^T, & x \in I^+ \\ (-1,1)^T, & x \in I^- \end{cases}$;
Step 4.	Construct the RDELM algorithm based classification equation (11) or (12);
Step 5.	Calculate the output vector for each image in the dataset. By Choosing the first value of output vector in the classifier $f(x)$ to output the similarity distance of the query. Besides, the score for each image I_i can be shown as $scores(I_i) = U^T f(x_i)$, $U = (1,0)^T$;
Step 6.	Sort the images by their score and return the new retrieval result.
Step 7.	Repeat Step2-Step6 until a satisfied results is obtained.

Sigmoid function is selected for ELM-based classification methods. The above process of the algorithm is obviously similar to the relevant feedback scheme based on SVM (Wang et al. 2011). However, this algorithm can be thousands times faster than the SVM based relevant feedback scheme and performs better while completing the same task. Its main reason is the utilizing of the ELM classifier, and our experiments illustrate the validity and time complexity of the algorithm.

5 Experimental results

In this section, we utilize four image retrieval datasets and three RF methods (SVM, ELM, MCVELM, RDELM and the RDELM-DRfree version methods) to evaluate the effect of our proposed RF image retrieval method. The four dataset are listed as follows: Corel-1K, Corel-5K, Corel-10K and MSRC image databases (see Figs. 6, 7, 8, 9, respectively). In following experiments, we re-write RDELM-DRfree method as DRfree for short.

In the following experiments, we assign $C = 0.05,\ \lambda = 1$. There are four RF times in our experiments, which are denoted by RF1, RF2, RF3 and RF4. The original image retrieval results mark as "Init". "FE method" represents feature extraction method. "STY1" indicates a feature extraction method, and includes global color histogram, LBP feature, and basis shape feature.

5.1 Image retrieval results

Corel-1K contains 10 categories and total 1000 images, includes bus, drag, food and so on. The size of each image is 256×384 or 384×256 (Fig. 6). The results of Corel-1K images are listed in Tables 1, 2, 3 and 4.

Fig. 6 Some samples of Corel-1K images

Fig. 7 Some samples of Corel-5K images

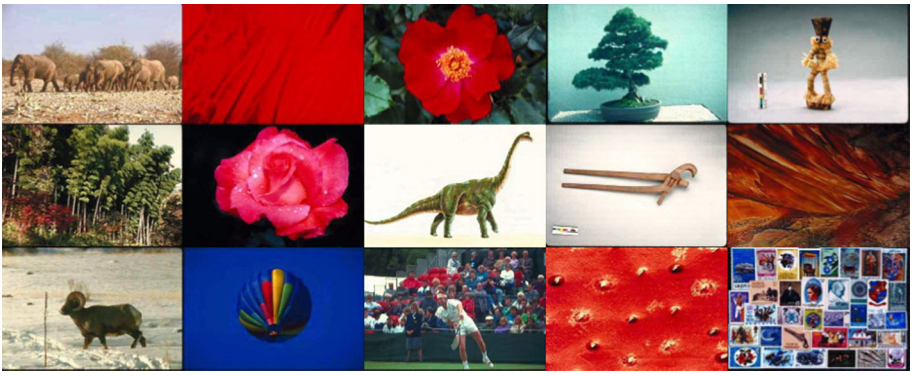

Fig. 8 Some samples of Corel-10K images

Fig. 9 Some samples of MSRC images

Table 1 The precision of STY1 and MSD features by SVM, ELM, MCVELM, DRfree and RDELM method in Corel-1K dataset (feedback 20 images) (%)

FE method	RF method	Init	RF1	RF2	RF3	RF4
STY1	SVM	52.4625	62.1722	65.8108	66.7117	67.4725
	ELM	52.4625	65.5856	74.2693	79.7798	83.7237
	RDELM	52.4625	69.049	75.3353	79.7548	82.5576
	DRFree	52.4625	67.5656	74.5312	79.7790	82.4505
	MCVELM	52.4625	64.8929	72.4675	78.0591	81.7107
MSD	SVM	67.1171	82.3674	85.7307	86.5916	87.022
	ELM	67.1171	81.967	85.4555	86.6466	87.3824
	RDELM	67.1171	82.2823	85.7508	87.047	87.8378
	DRfree	67.1171	81.8388	83.2382	85.2152	86.8519
	MCVELM	67.1171	81.0480	82.0971	83.7688	85.6056

Table 2 The precision of STY1 and MSD features by SVM, ELM, MCVELM, DRfree and RDELM method in Corel-1K dataset (feedback 30 images) (%)

FE method	RF method	Init	RF1	RF2	RF3	RF4
STY1	SVM	47.8545	60.6073	64.6847	65.5822	66.2863
	ELM	47.8545	60.7774	70.1735	76.5999	81.0177
	RDELM	47.8545	67.9947	74.2576	79.0023	82.1388
	DRFree	47.8545	65.7147	72.8942	77.1665	81.8782
	MCVELM	47.8545	63.6643	70.8051	76.4201	80.1985
MSD	SVM	62.9162	81.3146	84.4511	85.1318	85.5088
	ELM	62.9162	81.0577	84.1909	85.2219	85.7891
	RDELM	62.9162	81.1111	84.4745	85.4421	86.1828
	DRfree	62.9162	80.8178	82.5726	84.3410	86.1662
	MCVELM	62.9162	80.8345	82.2155	84.0574	85.5289

Table 3 The precision of STY1 and MSD features by SVM, ELM, MCVELM, DRfree and RDELM method in Corel-1K dataset (feedback 40 images) (%)

FE method	RF method	Init	RF1	RF2	RF3	RF4
STY1	SVM	44.452	59.0465	63.3208	63.9965	64.6947
	ELM	44.452	57.3649	66.8844	73.0531	78.1031
	RDELM	44.452	66.4289	72.8879	77.8654	80.6957
	DRFree	44.452	64.7673	70.5105	75.9880	79.6371
	MCVELM	44.452	62.6642	67.2187	74.6226	78.1742
MSD	SVM	59.4344	80.1176	82.9755	83.6737	84.0591
	ELM	59.4344	80.03	83.0305	84.0415	84.5796
	RDELM	59.4344	80.523	83.6461	84.532	85.1502
	DRfree	59.4344	79.9620	81.6992	83.1306	84.7372
	MCVELM	59.4344	80.0596	80.9284	82.3899	84.0265

Table 4 The precision of STY1 and MSD features by SVM, ELM, MCVELM, DRfree and RDELM method in Corel-1K dataset (feedback 50 images) (%)

FE method	RF method	Init	RF1	RF2	RF3	RF4
STY1	SVM	41.958	57.7277	61.7417	62.3564	63.1752
	ELM	41.958	54.8068	65.001	71.2412	76.3003
	RDELM	41.958	64.7708	71.2432	76.3443	79.3714
	DRFree	41.958	61.4805	67.2422	72.2322	77.6436
	MCVELM	41.958	60.1640	65.0861	70.2112	75.7187
MSD	SVM	56.7568	79.3714	82.1762	82.6987	83.1011
	ELM	56.7568	79.4094	82.0881	82.9049	83.5155
	RDELM	56.7568	79.7477	82.8048	83.5876	84.1141
	DRfree	56.7568	79.0581	80.8248	82.0901	83.6837
	MCVELM	56.7568	79.0961	80.2102	81.5155	82.9269

Table 5 The precision of STY1 and MSD features by SVM, ELM, MCVELM, DRfree and RDELM method in Corel-5K dataset (feedback 20 images) (%)

FE method	RF method	Init	RF1	RF2	RF3	RF4
STY1	SVM	24.775	27.115	29.1	29.93	30.35
	ELM	24.775	30.255	34.17	36.805	39.41
	RDELM	24.775	30.51	34.165	37.14	39.73
	DRFree	24.775	30.4900	34.1800	36.6850	39.0650
	MCVELM	24.775	30.0425	33.7548	36.6593	38.8320
MSD	SVM	27.215	38.345	42.195	43.84	45
	ELM	27.215	38.74	42.24	43.715	44.93
	RDELM	27.215	38.615	42.385	43.805	45.155
	DRfree	27.215	39.8000	40.4950	42.1250	43.5350
	MCVELM	27.215	38.9853	39.8626	41.1675	42.5778

Table 6 The precision of STY1 and MSD features by SVM, ELM, MCVELM, DRfree and RDELM method in Corel-5K dataset (feedback 30 images) (%)

FE method	RF method	Init	RF1	RF2	RF3	RF4
STY1	SVM	21.2	24.073	25.893	26.853	27.373
	ELM	21.2	26.61	30.23	32.897	35.367
	RDELM	21.2	27.023	30.81	33.677	35.987
	DRFree	21.2	26.1367	30.6567	32.2167	35.6100
	MCVELM	21.2	26.0110	30.2807	31.8116	35.3416
MSD	SVM	23.48	36.263	40.197	41.933	43.01
	ELM	23.48	36.49	40.173	41.67	42.633
	RDELM	23.48	36.677	40.12	41.633	42.773
	DRfree	23.48	38.1667	38.8900	40.5733	41.9733
	MCVELM	23.48	37.2609	38.7925	39.6084	41.4880

Corel-5K contains 50 categories and total 5000 images. The size of each image is 192×128 or 128×192 (Fig. 7). The results of Corel-5K images are listed in Tables 5, 6, 7 and 8.

The Corel-10K contains 100 categories and total 10000 images. The size of each image is 126×187 or 187×126 (Fig. 8). The results of Corel-10K images are listed in Tables 9, 10, 11 and 12.

We select a sub-dataset which includes 30 images for each of the 19 classes from MSRC dataset (as shown in Fig. 9). The size of each image is about 320×240. The results of MSRC images are listed in Tables 13, 14 and 15.

5.2 Parameters analysis

In this subsection, we discuss the influence of parameters in Eq. (9). Three parameters λ, C and $\tilde{N}$ are illustrated as follows. The parameter C is used to balance the Structural risk and empiric risk in Eq. (9). In image retrieval task, structural risk is more important than empiric risk because of small size of labeled samples. In our paper, we set $C = 0.05$ in all experiments

Table 7 The precision of STY1 and MSD features by SVM, ELM, MCVELM, DRfree and RDELM method in Corel-5K dataset (feedback 40 images) (%)

FE method	RF method	Init	RF1	RF2	RF3	RF4
STY1	SVM	19.005	22.23	23.99	24.865	25.223
	ELM	19.005	25.165	28.585	31.24	33.477
	RDELM	19.005	24.805	28.418	31.282	33.9
	DRFree	19.005	24.4425	28.8600	31.4825	33.6575
	MCVELM	19.005	23.9952	28.5217	31.2921	33.3349
MSD	SVM	21.488	35.063	39.27	40.847	41.763
	ELM	21.488	35.557	39.15	40.485	41.44
	RDELM	21.488	35.495	39.278	40.725	41.727
	DRfree	21.488	37.2975	37.8450	39.2875	40.4125
	MCVELM	21.488	36.1705	37.5665	39.1299	39.6122

Table 8 The precision of STY1 and MSD features by SVM, ELM, MCVELM, DRfree and RDELM method in Corel-5K dataset (feedback 50 images) (%)

FE method	RF method	Init	RF1	RF2	RF3	RF4
STY1	SVM	17.442	20.916	22.954	23.69	24.156
	ELM	17.442	23.78	27.102	29.802	31.976
	RDELM	17.442	23.594	27.032	29.856	32.208
	DRFree	17.442	23.6980	27.0460	29.7020	31.8840
	MCVELM	17.442	23.5312	26.6102	29.5329	31.7659
MSD	SVM	19.728	33.97	38.258	39.892	40.7
	ELM	19.728	34.63	38.33	39.574	40.474
	RDELM	19.728	34.49	38.134	39.52	40.518
DRfree	19.728	36.4360	37.0080	38.4200	39.4220	
MCVELM	19.728	35.5226	36.4611	37.4494	39.2801	

Table 9 The precision of STY1 and MSD features by SVM, ELM, MCVELM, DRfree and RDELM method in Corel-10K dataset (feedback 20 images) (%)

FE method	RF method	Init	RF1	RF2	RF3	RF4
STY1	SVM	23.9	25.832	26.844	27.268	27.564
	ELM	23.9	27.12	30.924	33.292	35.68
	RDELM	23.9	27.116	30.856	33.26	35.684
	DRfree	23.9	27.3600	30.5200	33.9640	35.8920
	MCVELM	23.9	26.5798	30.1143	33.3047	35.7772
MSD	SVM	37.46	46.616	48.732	49.564	49.968
	ELM	37.46	45.524	47.632	48.468	49.212
	RDELM	37.46	45.532	47.604	48.524	49.22
	DRfree	37.46	46.0168	47.3788	48.7322	49.2117
	MCVELM	37.46	45.3332	47.1810	48.5792	49.0966

Table 10 The precision of STY1 and MSD features by SVM, ELM, MCVELM, DRfree and RDELM method in Corel-10K dataset (feedback 30 images) (%)

FE method	RF method	Init	RF1	RF2	RF3	RF4
STY1	SVM	20.168	22.496	23.408	23.797	24.016
	ELM	20.168	24.123	27.763	30.435	32.632
	RDELM	20.168	23.971	27.571	30.075	32.536
	DRfree	20.168	24.2053	28.6347	30.9333	32.6933
	MCVELM	20.168	23.4001	28.0961	30.4345	32.2548
MSD	SVM	32.648	43.64	45.955	46.795	47.339
	ELM	32.648	42.251	44.28	45.307	46.053
	RDELM	32.648	42.517	44.653	45.563	46.315
	DRfree	32.648	42.5498	43.9608	44.9072	46.1786
	MCVELM	32.648	42.4671	43.5410	44.4693	45.2294

Table 11 The precision of STY1 and MSD features by SVM, ELM, MCVELM, DRfree and RDELM method in Corel-10K dataset (feedback 40 images) (%)

FE method	RF method	Init	RF1	RF2	RF3	RF4
STY1	SVM	17.854	20.736	21.66	22.024	22.288
	ELM	17.854	22.256	25.866	28.47	30.486
	RDELM	17.854	22	25.568	28.152	30.458
	DRfree	17.854	22.9340	25.8060	28.5940	30.3420
	MCVELM	17.854	22.8937	25.7948	27.6845	30.1959
MSD	SVM	29.51	41.572	44.108	44.982	45.488
	ELM	29.51	40.398	42.972	43.826	44.44
	RDELM	29.51	40.594	42.818	43.656	44.292
	DRfree	29.51	40.4171	42.9217	43.1185	43.7399
	MCVELM	29.51	40.3783	42.2788	42.7534	43.3636

Table 12 The precision of STY1 and MSD features by SVM, ELM, MCVELM, DRfree and RDELM method in Corel-10K dataset (feedback 50 images) (%)

FE method	RF method	Init	RF1	RF2	RF3	RF4
STY1	SVM	16.102	19.174	20.122	20.509	20.715
	ELM	16.102	20.883	24.254	26.749	28.755
	RDELM	16.102	20.083	23.547	26.198	28.41
	DRfree	16.102	20.9648	23.2560	26.5504	28.3040
	MCVELM	16.102	20.7011	22.5595	25.9867	28.2543
MSD	SVM	27.11	39.728	42.683	43.726	44.23
	ELM	27.11	38.811	41.261	42.166	42.846
	RDELM	27.11	39.046	41.451	42.293	42.866
	DRfree	27.11	38.9424	40.9541	42.2439	42.7135
	MCVELM	27.11	38.0825	40.0968	42.2002	42.8174

Table 13 The precision of STY1 and MSD features by SVM, ELM, MCVELM, DRfree and RDELM method in MSRC dataset (feedback 10 images) (%)

FE method	RF method	Init	RF1	RF2	RF3	RF4
STY1	SVM	46.526	58.018	60.982	61.807	61.982
	ELM	46.526	57.772	60.754	61.702	61.982
	RDELM	46.526	57.895	61.018	61.947	62.228
	DRfree	46.526	60.2427	61.2842	61.8246	
	MCVELM	46.526	59.6562	60.7716	61.7690	
MSD	SVM	46.526	57.877	60.947	61.842	62.07
	ELM	46.526	57.737	60.596	61.561	61.807
	RDELM	46.526	57.947	60.965	61.86	62.088
	DRfree	46.526	58.0175	60.8070	61.4912	62.0386
	MCVELM	46.526	57.6288	59.8735	60.6023	62.0673

Table 14 The precision of STY1 and MSD features by SVM, ELM, MCVELM, DRfree and RDELM method in MSRC dataset (feedback 20 images) (%)

FE method	RF method	Init	RF1	RF2	RF3	RF4
STY1	SVM	37.237	50.132	54.64	56.289	56.807
	ELM	37.237	49.798	54.114	55.851	56.579
	RDELM	37.237	49.965	54.43	56.246	56.947
	DRfree	37.237	54.1404	56.1281	56.8070	
	MCVELM	37.237	53.6643	55.6138	55.8595	
MSD	SVM	37.237	50.123	54.64	56.325	56.886
	ELM	37.237	49.658	54.123	55.842	56.526
	RDELM	37.237	50.123	54.649	56.325	56.895
	DRfree	37.237	50.0439	54.8509	55.8772	56.8982
	MCVELM	37.237	50.0336	54.3829	54.9697	56.7576

Table 15 The precision of STY1 and MSD features by SVM, ELM, MCVELM, DRfree and RDELM method in MSRC dataset (feedback 30 images) (%)

FE method	RF method	Init	RF1	RF2	RF3	RF4
STY1	SVM	32.83	45.772	50.474	52.532	53.351
	ELM	32.83	45.24	49.965	52.053	53.094
	RDELM	32.83	45.497	50.211	52.298	53.263
	DRfree	32.83	45.561	49.4298	52.4298	53.2982
	MCVELM	32.83	44.925	49.1126	51.8077	52.7909
MSD	SVM	32.83	45.76	50.444	52.491	53.339
	ELM	32.83	45.24	49.86	51.901	52.743
	RDELM	32.83	45.678	50.386	52.427	53.257
	DRfree	32.83	45.7485	50.4620	52.5263	53.1111
	MCVELM	32.83	44.8621	49.8403	51.9922	52.7122

and data sets. We involved λ to avoid the optimization matrix irreversible in Eq. (10). We set $\lambda = 1$ in our method. In fact, λ can be set by $\lambda = tr(S'_w - S'_b)$ in any cases. In ELM, the parameter $\tilde{N}$ is also important, more details can be seen in references Huang et al. (2006) and Huang and Chen (2007). We set $\tilde{N} = 200$ in all ELM versions and experiments.

5.3 Results analysis

RDELM and the DRfree version achieve better performance than SVM, MCVELM and ELM in most cases, especially in Corel-1K dataset (see Tables 1, 2, 3, 4). This is mainly because the classes of images keep good structures in RDELM space. There is also an interesting result that retrieval results of MSD are always better than that of STY1 (see Tables 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12). It illustrates that a good feature is very important, and needs fewer times in RF. However, In MSRC dataset, SVM, ELM, MCVELM and RDELM achieved the similar results which illustrate the classes in ELM space and original sample space are not distinguishable. This can be also explained by the similar results using STY1 or MSD features (see Tables 13, 14, 15).

RDELM and the DRfree version get better performance than SVM, MCVELM and ELM in Corel-1K dataset, which can be also attributed to the robust discriminative structure of RDELM. Corel-1K is more difficult to retrieve than other datasets. For example, the classes of beach and mountain share some similar features in Corel-1K dataset. From the above experimental results in four datasets, we claim that a good descriptor of image may be discriminative which can also help RF and the discriminative structure in score model.

The efficiency of RDELM is similar to MCVELM because both the two methods need dimensionality reduction with the similar process and dimension. The time complexity of dimensionality reduction in RDELM needs $O(D^3)$. This is an extra time consuming to ELM. Therefore, RDELM need more time to get the feedback function. However, it makes little influence while the feature dimensions of image are not very high (more than 500).

6 Conclusion

In this paper, we deeply analyzed the class structures in ELM feature space, and propose a robust ELM version for RF in CBIR system (RDELM). RDELM can select whether the dimensionality reduction method need to be involved or not, and is suitable for image retrieval application. Experimental results on benchmark datasets show that ELM-related methods are more efficient than SVM-based method. Furthermore, the results of RF1 illustrate image feature extraction plays an important role in image retrieval process.

References

Akusok, A., Miche, Y., Karhunen, J., et al. (2015). Arbitrary category classification of websites based on image content. *IEEE on Computational Intelligence Magazine*, *10*(2), 30–41.

Anitha, S., & Rinesh, S. (2013). Semi-supervised biased maximum margin analysis for interactive image retrieval. *Research Journal of Computer Systems Engineering*, *4*, 532–536.

Cao, J., Huang, W., Zhao, T., Wang, J., & Wang, R. (2015a). An enhance excavation equipments classification algorithm based on acoustic spectrum dynamic feature. *Multidimensional Systems and Signal Processing*. doi:10.1007/s11045-015-0374-z.

Cao, J., & Lin, Z. (2015). Extreme learning machine on high dimensional and large data applications: A survey. *Mathematical Problems in Engineering*. doi:10.1155/2015/103796.

Cao, J., Lin, Z., Huang, G.-B., & Liu, N. (2012). Voting based extreme learning machine. *Information Sciences*, *185*(1), 66–77.

Cao, J., Zhao, Y., Lai, X., Ong, M., Yin, C., Koh, Z., et al. (2015b). Landmark recognition with sparse representation classification and extreme learning machine. *Journal of The Franklin Institute*, *352*(10), 4528–4545.

Deng, W., Zheng, Q., & Chen, L. (2009). Regularized extreme learning machine. In *Computational intelligence and data mining, CIDM'09* (pp. 389–395).

Feng, L., Liu, S., Xiao, Y., et al. (2015). A novel CBIR system with WLLTSA and ULRGA. *Neurocomputing*, *147*, 509–522.

He, X. (2004). Incremental semi-supervised subspace learning for image retrieval. In *Proceedings of the 12th annual ACM international conference on multimedia* (pp. 2–8).

He, X., & Niyogi, P. (2003). Locality preserving projections. In *Advances in neural information processing systems 16*. Vancouver, Canada.

He, Q., Jin, X., Du, C., et al. (2014). Clustering in extreme learning machine feature space. *Neurocomputing*, *128*, 88–95.

Hoi, S. C. H., Jin, R., Zhu, J., et al. (2008) Semi-supervised SVM batch mode active learning for image retrieval. In *IEEE Conference on Computer Vision and Pattern Recognition* (pp. 1–7).

Hoi, S. C. H., & Lyu, M. R. (2005). A semi-supervised active learning framework for image retrieval. *Computer Vision and Pattern Recognition*, *2*, 302–309.

Horata, P., Chiewchanwattana, S., & Sunat, K. (2013). Robust extreme learning machine. *Neurocomputing*, *102*, 31–44.

Huang, G.-B. (2015). What are extreme learning machines? Filling the gap between Frank Rosenblatt's Dream and John von Neumann's Puzzle. *Cognitive Computation*, *7*, 263–278.

Huang, G., & Chen, L. (2007). Convex incremental extreme learning machine. *Neurocomputing*, *70*(16–18), 3056–3062.

Huang, G., Chen, L., & Siew, C.-K. (2006). Universal approximation using incremental constructive feedforward networks with random hidden nodes. *IEEE Transactions on Neural Networks*, *17*(4), 879–892.

Huang, G. B., Zhou, H., Ding, X., et al. (2012). Extreme learning machine for regression and multiclass classification. *IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics*, *42*(2), 513–529.

Iosifidis, A., Tefas, A., & Pitas, I. (2013). Minimum class variance extreme learning machine for human action recognition. *IEEE Transactions on Circuits and Systems for Video Technology*, *23*(11), 1968–1979.

Iosifidis, A., Tefas, A., & Pitas, I. (2014). Regularized extreme learning machine for multi-view semi-supervised action recognition. *Neurocomputing*, *145*, 250–262.

Jin, Y., Cao, J., Wang, Y., et al. (2015). Ensemble based extreme learning machine for cross-modality face matching. *Multimedia Tools and Applications*, 1–16.

Kundu, M. K., Chowdhury, M., & Bulò, S. R. (2015). A graph-based relevance feedback mechanism in content-based image retrieval. *Knowledge-Based Systems*, *73*, 254–264.

Liu, S., Feng, L., & Qiao, H. (2015). Scatter Balance: An angle-based supervised dimensionality reduction. *IEEE Transactions on Neural Networks and Learning Systems*, *26*(2), 277–289.

Liu, S., Feng, L., Xiao, Y., et al. (2014). Robust activation function and its application: Semi-supervised kernel extreme learning method. *Neurocomputing*, *144*, 318–328.

Liu, G. H., Li, Z. Y., Zhang, L., et al. (2011). Image retrieval based on micro-structure descriptor. *Pattern Recognition*, *44*(9), 2123–2133.

Lu, K., Zhao, J., & Cai, D. (2006). An algorithm for semi-supervised learning in image retrieval. *Pattern Recogition*, *39*(4), 717–720.

Minhas, R., Baradarani, A., Seifzadeh, S., et al. (2010). Human action recognition using extreme learning machine based on visual vocabularies. *Neurocomputing*, *73*(10), 1906–1917.

Mohammed, A. A., Minhas, R., Wu, Q. M. J., et al. (2011). Human face recognition based on multidimensional PCA and extreme learning machine. *Pattern Recognition*, *44*(10), 2588–2597.

Murala, S., & Wu, Q. M. (2014). Local mesh patterns versus local binary patterns: Biomedical image indexing and retrieval. *IEEE Journal of Biomedical and Health Informatics*, *18*(3), 929–938.

Ojala, T., Pietikainen, M., & Maenpaa, T. (2002). Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *24*(7), 971–987.

Swain, M. J., & Ballard, D. H. (1991). Color indexing. *International Journal of Computer Vision*, *7*(1), 11–32.

Tan, X., & Triggs, B. (2007). Enhanced local texture feature sets for face recognition under difficult lighting conditions. In *Analysis and modeling of faces and gestures* (pp. 168–182). Berlin: Springer.

Tang, X., & Han, M. (2009). Partial Lanczos extreme learning machine for single-output regression problems. *Neurocomputing*, *72*(13–15), 3066–3076.

Tang, X., & Han, M. (2009). Partial Lanczos extreme learning machine for single-output regression problems. *Neurocomputing*, *72*(13–15), 3066–3076.

Zhang, P., & Yang, Z. (2015). A robust AdaBoost.RT based ensemble extreme learning machine. *Mathematical Problems in Engineering, 2015,* 260970. http://www.hindawi.com/journals/mpe/2015/260970/cta/.

Zhang, K., & Luo, M. (2015). Outlier-robust extreme learning machine for regression problems. *Neurocomputing*, *151*, 1519–1527.

Zhang, S., Yang, M., Cour, T., et al. (2015). Query specific rank fusion for image retrieval. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 37*(4), 803–815.