

Leukocyte image segmentation by visual attention and extreme learning machine

Chen Pan · Dong Sun Park · Yong Yang ·
Hyouck Min Yoo

Received: 13 September 2010 / Accepted: 16 December 2010 / Published online: 30 January 2011
© Springer-Verlag London Limited 2011

Abstract This paper presents a fast and simple framework for leukocyte image segmentation by learning with extreme learning machine (ELM) and sampling via simulating visual system. In sampling stage, visual attention and the effect of microsaccades in fixation are simulated. The high gradient pixels in fixation regions are sampled to group training set. We designed an automatic sampling process for leukocyte image according to the staining knowledge of blood smears. In learning stage, ELM classifier is trained online to simulate visual neuron system and then extracts pixels of object from image. The ELM-based segmentation is fully automatic by the proposed framework, which could find efficient samples actively, train the classification model in real time and almost no parameter adjusted. Experimental results demonstrated the new method could extract entire leukocyte from complex scenes, has equivalent performance compared to the SVM-based method and exceeds the marker-controlled watershed algorithm.

Keywords Image segmentation · Visual attention · Extreme learning machine · Real-time learning · Leukocyte

C. Pan (✉)
College of Information Engineering, China Jiliang University,
310018 Hangzhou, Zhejiang, China
e-mail: pc916@cjlu.edu.cn

C. Pan · D. S. Park (✉) · H. M. Yoo
Division of Electronics and Information Engineering,
Chonbuk National University, Jeonju, Jeonbuk 561-756, Korea
e-mail: dspark@jbnu.ac.kr

Y. Yang
School of Information Technology,
Jiangxi University of Finance and Economics,
330013 Nanchang, Jiangxi, China

1 Introduction

Microscopic leukocyte analysis is a powerful tool for diagnosing many types of diseases. Computer-aided automatic analysis could enhance the objectivity of the diagnosis, save manpower, and time. However, the complex biological nature of the leukocyte and the technical problems which are caused by unstandardized smear preparation and image acquisition often make the cell image with complex color and texture. It is a challenge to segment entire leukocyte due to the feature of cell may be uncertain in dynamic scene. It is an unsolved issue in blood cell image segmentation. This work is important in an area of medical image analysis as part of an automatic diagnostic method for certain medical conditions/diseases, e.g. leukemia. There is a large literature on cell image segmentation. A classical approach is the marker-controlled watershed method from morphology [2], where the watershed lines are computed on a gradient-based topographic surface obtained by imposing a selected set of markers as only regional minima. However, the ideal markers usually need be input manually by human. It is difficult to locate suitable markers automatically due to complexity of cell image.

Natural images are typical unstructured data. Such data may have high-dimensional features, may be incomplete and uncertain in content, hardly to characterize with limited rules, and interpretation usually depends on user. How to create a computational model for such data is a hot research field of artificial intelligence. Learning by sampling from data is an effective strategy to modeling unstructured data. However, how to select the learning algorithm and where to sample from the data become very critical. Since those factors determine the performance of the model in practice.

In past decade, neural network (NN) and support vector machine (SVM) have been successfully applied to cell image segmentation. Wang and Wang [17] presented an algorithm based on fuzzy cellular neural networks (FCNN) to detect leukocyte. Pan et al. [14] use mean-shift algorithm and SVM to segment leukocyte image. Both of them have excellent ability of nonlinear approximation to provide models that are difficult to handle using classical parametric techniques. However, besides training samples, supervised learning needs tune/set key parameters carefully, such as the controlling parameters of FCNN in [17], the width of the search window of mean-shift and the kernel parameters of SVM in [14]. That is not an easy job and time-consuming. In conventional supervised image segmentation algorithms, training is seldom in real time due to time cost so that few supervised algorithm can generate segmentation model online adaptively. Most of supervised algorithms use a number of images training off-line and then produce a model to segment other images [18], since the model is fixed that is not good to deal with the uncertainty and changing that often faced in natural images.

Recently, Huang et al. [6, 7] proposed a novel machine learning algorithm namely extreme learning machine (ELM) that can significantly reduce the training time of an NN. The ELM theory shows that the hidden nodes of the “generalized” single-hidden layer feedforward networks (SLFNs), which need not be neuron alike, can be randomly generated and the universal approximation capability of such SLFNs can be guaranteed. ELM can analytically determine all the parameters of SLFNs instead of adjusting parameters iteratively. The latest research [7] shows that (1) SVM’s maximal margin property and the minimal norm of weights theory of feedforward neural networks are consistent actually under the ELM learning framework; (2) in classification, ELM and SVM are equivalent when the standard optimization method is used to them, but ELM has less optimization constraints due to any set of distinct training data transformed from the input space to the ELM feature space with the activation function are linearly separable. ELM in classification tends to achieve better generalization performance than traditional SVM, less sensitive to user specified parameters, and could be implemented easily.

As we know, the capacity of the human vision processing information far exceeds the current machine vision. Modern research from psychophysical and neurophysiological experiments have found that the primate visual system employs an attention mechanism to limit processing to important information that is currently relevant to behaviors or visual tasks. It can efficiently deal with the balance between computing resources, time cost and performing different visual tasks in a normal, cluttered and

dynamic environment [16]. So simulate human visual attention may be an effective approach to sampling for ELM.

In this paper, we propose a novel two-stage method for complex image segmentation. In sampling stage, we firstly locate the regions of interesting (ROI) according to the special color, and then dilate the ROI in a rule to enhance the entropy of the region continually. Over-sampling and resampling could be considered in our method in order to get more accurate segmentation. In learning stage, ELM classifier is trained online and extracts objects from the image. Experimental results in color cell images demonstrated that the new method has better performance compared to the watershed-based and SVM-based methods in complex scenes.

This paper is organized as following. Section 2 briefly introduces ELM and sampling strategy in segmentation. Section 3 shows the framework of the method and application in color leukocyte image segmentation. Section 4 compares experimental results with watershed-based and SVM-based methods and gives some discussion. Conclusion is in Sect. 5.

2 Theory and method

2.1 Brief of ELM

ELM is a unified SLFN with randomly generated hidden nodes independent of the training data [4, 6, 7]. For N arbitrary distinct samples $(\mathbf{x}_i, \mathbf{t}_i)$, where $\mathbf{x}_i = [x_{i1}, x_{i2}, \dots, x_{in}]^T \in R^n$ and $\mathbf{t}_i = [t_{i1}, t_{i2}, \dots, t_{im}]^T \in R^m$ (n is the number of dimensions of input \mathbf{x} , m is the number of classes of data). So a given set of training samples $\{(\mathbf{x}_i, \mathbf{t}_i)\}_{i=1}^N \subset R^n \times R^m$, the output of a SLFN with L hidden nodes can be represented by

$$f_L(\mathbf{x}_j) = \sum_{i=1}^L \beta_i K(\boldsymbol{\alpha}_i, b_i, \mathbf{x}_j) = \mathbf{t}_j, \quad j = 1, \dots, N \quad (1)$$

where $\boldsymbol{\alpha}_i$ and b_i are the parameters of hidden node which could be randomly generated. $K(\boldsymbol{\alpha}_i, b_i, \mathbf{x})$ is the output of the i th hidden node with respect to the input \mathbf{x} . And β_i is the weight connecting the i th hidden node to the output node. Equation (1) can be written compactly as

$$\mathbf{H}\boldsymbol{\beta} = \mathbf{T} \quad (2)$$

where

$$\mathbf{H}(\boldsymbol{\alpha}_1, \dots, \boldsymbol{\alpha}_L, b_1, \dots, b_L, \mathbf{x}_1, \dots, \mathbf{x}_N) = \begin{bmatrix} K(\boldsymbol{\alpha}_1, b_1, \mathbf{x}_1) & \cdots & K(\boldsymbol{\alpha}_L, b_L, \mathbf{x}_1) \\ \vdots & \cdots & \vdots \\ K(\boldsymbol{\alpha}_1, b_1, \mathbf{x}_N) & \cdots & K(\boldsymbol{\alpha}_L, b_L, \mathbf{x}_N) \end{bmatrix}_{N \times L} \quad (3)$$

$$\beta = \begin{bmatrix} \beta_1^T \\ \vdots \\ \beta_L^T \end{bmatrix}_{L \times m} \quad \text{and} \quad \mathbf{T} = \begin{bmatrix} \mathbf{t}_1^T \\ \vdots \\ \mathbf{t}_N^T \end{bmatrix}_{N \times m} \tag{4}$$

β^T is the transpose of a matrix or vector β . \mathbf{H} is called the hidden layer output matrix of the network [6]; the i th column of \mathbf{H} is the i th hidden node’s output vector with respect to inputs $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_N$. and the j th row of \mathbf{H} is the output vector of the hidden layer with respect to input \mathbf{x}_j . It has been proved in theory [4, 6] that SLFNs with random hidden nodes have the universal approximation capability, and the hidden nodes can be randomly generated independent of the training data.

After the hidden nodes are randomly generated and given the training data, the hidden-layer output matrix \mathbf{H} is known and need not be tuned. Thus, training SLFNs simply amounts to getting the solution of a linear system (2) of output weights β .

According to Bartlett’s theory [1] for feedforward neural networks, in order to get the better generalization performance, ELM not only tries to reach the smallest training error but also the smallest norm of output weights.

Minimize: $\|\mathbf{H}\beta - \mathbf{T}\|$

and

Minimize: $\|\beta\|$ (5)

In the case of binary classification, Huang et al. [7] proved that to minimize the norm of the output weights $\|\beta\|$ is actually to maximize the distance of the separating margins of the two different classes $2/\|\beta\|$ in ELM feature space. Under the constraint of equation (5), a simple representation of the solution of the system (2) is given explicitly by Huang et al. [6] as

$$\hat{\beta} = \mathbf{H}^\dagger \mathbf{T} \tag{6}$$

where \mathbf{H}^\dagger is the Moore–Penrose generalized inverse of the hidden-layer output matrix \mathbf{H} .

If the N training data are distinct, \mathbf{H} is column full rank with probability one when $L \leq N$. In real applications, the number of hidden nodes is always less than the number of training data $L < N$. Thus

$$\mathbf{H}^\dagger = (\mathbf{H}^T \mathbf{H})^{-1} \mathbf{H}^T \tag{7}$$

Huang et al. [6, 7] have proved SLFNs with a wide type of random computational hidden nodes. Additive and RBF hidden nodes are used often in applications. For example, additive hidden node with the activation function $k(\mathbf{x}) : R \rightarrow R$ (e.g., sigmoid, threshold, sin/cos, etc.), $K(\alpha_i, b_i, \mathbf{x})$ is given by

$$K(\alpha_i, b_i, \mathbf{x}) = k(\alpha_i \cdot \mathbf{x} + b_i) \tag{8}$$

where α_i is the weight vector connecting the input layer to the i th hidden node and b_i is the bias of the i th hidden node. $\alpha_i \cdot \mathbf{x}$ denotes the inner product of vectors α_i and \mathbf{x} in R^n . The three-step simple learning algorithm can be summarized as follows

Algorithm 1: Learning of ELM

- Given a training set $\{(\mathbf{x}_i, \mathbf{t}_i)\}_{i=1}^N \subset R^n \times R^m$, the hidden-node output function $K(\alpha_i, b_i, \mathbf{x})$, and hidden-node number L :
1. Randomly assign hidden-node parameters $(\alpha_i, b_i), i = 1, \dots, L$;
 2. Calculate the hidden-layer output matrix \mathbf{H} ;
 3. Calculate the output weight vector $\beta : \beta = \mathbf{H}^\dagger \mathbf{T}$.

2.2 Strategy for sampling

Eye is a natural sampling system. Most of the time, our eyes scan visual scene in sequences of saccades and fixations [12]. Saccades aim for visual information currently outside the fovea of the retina. Fixation keep a target relatively stable with respect to the photoreceptors on the retina. Notably, during fixation, our eyes move continuously rather than holding steady. The very small, involuntary flick in eye position is called microsaccades. Although the precise nature of visual perception remains unclear, it is generally agreed that eye movements are very important in vision. The spatiotemporal characteristics of saccades and microsaccades may reflect an optimal sampling method by which the brain discretely acquires visual information [8].

Microsaccades can move a stationary stimulus in and out of a neuron’s receptive field (like trembling), thereby producing transient neural responses to reduce perceptual fading and keep the continuity of perception. Thus, spatial discontinuity in visual field would cause dynamic stimulus result in the excitability of neurons by microsaccades. The uniform regions with zero gradients may be stationary stimulus and will fade during visual fixation.

We suppose that the most important information is the pixels on edge, but the complex images always have various edges and most of them may be unnecessary details and noise. How to locate the effective edges for image task is a problem.

Shannon entropy of local attributes is often used to define saliency in terms of local signal complexity or unpredictability [9]. Given an original region, a gradient threshold s is used to select high gradient pixels to form a new region $R(s)$, and a descriptor D that takes on values $\{d_1, \dots, d_r\}$ (e.g. in an 8 bit grey level image D would range from 0 to 255), local entropy is defined as a function of s :

$$E_{D,R}(s) = - \sum_i p_{D,R}(s, d_i) \log_2 p_{D,R}(s, d_i) \quad (9)$$

where $p_{D,R}(s, d_i)$ is the probability of descriptor D taking the value d_i in the local region R with s gradient level (or scale).

We note that the entropy of new region is always higher than that of the original region in some gradient level according to formula (9). It means that edges could have higher “information content” in some scale to form a peak of entropy. Figure 1 gave an instance. So the aim of sampling in our method is the high gradient pixels in the region with the maximum entropy.

2.2.1 Sampling rule 1

In order to determine where edge (high gradient pixels) should be sampled, the best gradient threshold is

$$s = \arg \max E_{D,R}(s) \quad (10)$$

Color space is a natural feature space where each color is represented by a single point/vector. All the pixels with same color in an image will be mapped into a point in color space. Although this one-to-many mapping seemly lose spatial information of pixels, one of the benefits is that even if one pixel was known, all the pixels with same color in an image will be detected soon. The known vectors could be form a color look-up-table (CLUT) to segment corresponding pixels.

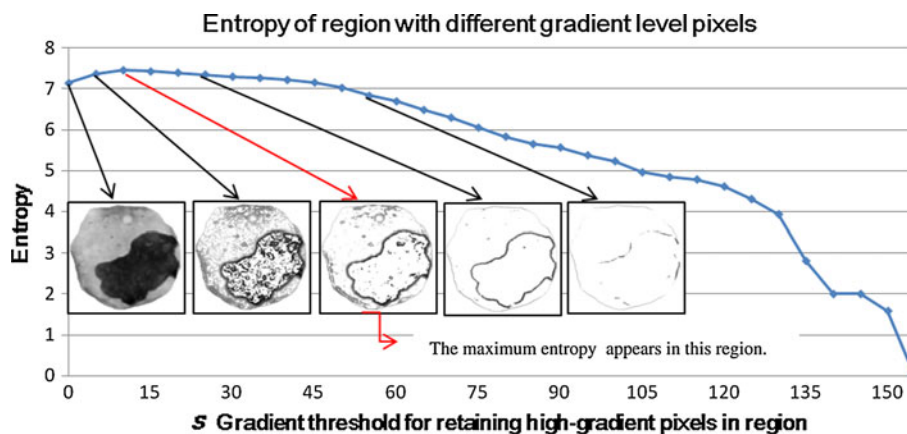
If a multicolor object contains subregions each with one color, the mapping vectors of the object could be represented by sampling from the pixels on the adjacent edges between those subregions within the object.

2.2.2 Sampling rule 2

In order to determine the number of sampling, the Nyquist–Shannon sampling theorem is applicable in our method.

$$N_{\text{sample}} > 2C_{\text{object}} \quad (11)$$

Fig. 1 This is a curve of entropy of a region formed with different gradient level pixels, where the low gradient pixels had been removed according to a gradient threshold. The maximum entropy appears in edge region with a suitable gradient level (or scale) rather than original region



where N_{sample} is the number of sampling from an object, and C_{object} is the number of color within the object, which is regarded as the frequency of color of the object.

Obviously, the adjacent edges within the object show discontinuity of color level on which pixel with higher gradient attracts our attention to focus on them. So image segmentation model will be constructed according to following idea.

To segment entire multicolor object, we firstly uniform sample N_{sample} pixels on the edges within the object regions as references, and then construct a classification model with ELM to make a CLUT in RGB color space for image segmentation.

2.3 Find local regions with saliency

In previous work, Fergus et al. [5] and Kadir and Brady [9] utilized an entropy-based method to find regions that are salient over both location and scale. For each point on the image, a histogram $P(L)$ is made of the intensities in a circular region of radius r . The entropy $H(r)$ of this histogram is then calculated, and the local maxima of $H(r)$ are candidate scales for the region. The saliency of each of these candidates is measured by $H \frac{dP}{dr}$. The M regions with highest saliency over the image could provide the features for learning and recognition. The saliency measure is invariant to scaling and could give stable identification of features. However, since above operation involves every pixel, it is very time-consuming in running. Moreover, this method just only uses monochrome information in spatial domain and does not take account of color and prior knowledge. In this paper, we improved and simplified entropy-based method using color, gradient, and prior knowledge.

Two types of visual attention exist in human vision system [16]. One is “Bottom-up”, in which low-level visually salient features are mostly used to attract visual attention. The other is “Top-down”, by which some objects/regions could be fixated in order to view details.

Information in bottom-up attention includes basic features such as color, orientation, motion, depth, conjunctions of features such as objects in 2D or 3D space [16]. A great number of models make use of “saliency” to direct attention. Saliency could be expressed in the feature space or the spatial domain. For examples, feature-based model detects salient clusters in color space, and space-based model always selecting continuous spatial areas in spatial domain.

The feature-based model could be simulated with many unsupervised algorithms that can detect the significant peaks of the probability density in feature space, such as thresholding, clustering (FCM, mean-shift) in color space. In our method, histogram analysis is utilized as a feature-based attention model to find significant peaks in color space, since it is simple and fast. Similarly [10], 1D histogram thresholding could be firstly applied to H, S, I components, respectively. Otsu’ algorithm [13] is used to locate a threshold in each color component’s histogram. Then, these thresholds are used to partition the color space into several hexahedra, one includes a class. So the image could be roughly segmented into several subregions with homogeneous color. In this step, we should ensure the main object with low under-segmentation error that means the segmented region contains pure pixels that belong to the object with less noise. Whereas high under-segmentation error means some noise exist. If object region were under-segmented seriously in this step, we need to increase the different component of color in order to over-segment the main object into many blocks each with similar color. For instance, the $I_1I_2I_3$ color system $I_1 = \frac{1}{3}(R + G + B)$, $I_2 = \frac{1}{2}(R - B)$, $I_3 = \frac{1}{4}(2G - R - B)$ may be more useful to descript image with difference of color. Because serious over-segmentation could reduce under-segmentation error effectively.

The space-based model is similar as region-growing or dilation in mathematical morphology, which extends from initial/seed region by selecting continuous spatial area according to some rule of adjacent pixels. Here, we adopt a conditional dilation to simulate space-based model. The dilation begins from an initial region, accepts adjacent pixel with an entropy increase rule iteratively. That is to say, if a candidate adjacent pixel could enhance the entropy of the region, it will become a new pixel of the region. Otherwise, the pixel will be ignored. By this way, the entropy of the dilated regions could be increased continually so that it achieves maximum soon. The dilation iteration will not stop until the entropy value of the region trends to decrease.

From the sampling point of view, feature-based model firstly provides the initial location of object for fixation. Then space-based model collects salient pixels around the initial location according to the entropy enhancement rule in order to achieve the most salient region and find the effective edges of object.

2.4 Sampling, training and testing

Considering image segmentation as a two-class classification problem, we sample the high gradient pixels uniformly from the object and non-object regions to group two-class samples for ELM training and then segment image using ELM model. The details, see Algorithm II.

It is worth mentioning that the color number of object could be determined automatically by hue histogram of object in a resolution (such as 64, 128, or 256 levels of color). The size of training set could be very small. For example, if color number of object do not exceed 30 in a resolution of hue 64, after 10 times sampling, the size of training set $N < 600$. So the training of ELM is nearly in real time.

The number of hidden nodes of ELM is the only factor that need be set by user. Huang et al. [6, 7] proved and demonstrated if the number of hidden nodes of ELM is large enough, ELM always tends to minimize the training error as well as the norm of the output weights (Convergence theorem in [6]). That means we can control segmentation accuracy by the number of hidden nodes. The large number of hidden nodes means the more computation complexity of ELM. We can choose it by trial-and-error or adopt error minimized ELM [6]. To simplify, an experience selection of the number of hidden nodes is one-thirtieth of the size of training set in 5 times sampling rate in this paper (when we adopt higher sampling rate in experiments, the number of hidden nodes need not change since the color number of objects does not change).

Algorithm II: Sampling, training and testing

1. Assume the size of training set is N .

2. Equivalent sampling $\frac{N}{2}$ pixels from positive and negative regions respectively. Positive pixels are marked with I^+ , and negative pixels are marked with I^- . In order to avoid uncertainty, we set $I^+ \cap I^- = \Phi$ (empty set). Let the samples $(\mathbf{x}_i, \mathbf{t}_i)$, $\mathbf{x}_i \in I^+ \cup I^-$,

$$\mathbf{t}_i = \begin{cases} (1, 0), & \text{if } \mathbf{x}_i \in I^+ \\ (0, 1), & \text{if } \mathbf{x}_i \in I^- \end{cases}, \mathbf{x}_i = (R, G, B)^T, \text{ where } (R, G, B) \text{ is the color value of pixel.}$$

3. Set the number of hidden nodes of ELM (int) $N/30$ in 5 times sampling rate.

4. Train an ELM online using the training set, ‘sigmoid’ activation function. Then generate a classifier model, which could be regard as a CLUT.

5. Use above model to classify the image pixels represented by $(R, G, B)^T$.

2.5 Over-sampling and resampling in visual task

In “top-down” attention, salient regions could be fixated. Gilchrist and Otero-Millan et al. [8] proposed that the dynamics of saccades and microsaccades may reflect an optimal strategy by which visual neurons discretely sample

information from a scene. Previous researchers have found microsaccades to be more prominent in conditions that involved identified targets and increased attentional demands [12]. Since both the area of retina and the number of sensors on retina are limited, the trembling of eyes results from microsaccades and saccades may achieve a mechanism to over-sampling and resampling.

Over-sampling means drawing repeated samples from the given data. It is the basis of image super-resolution (SR) processing. The latter is the technology to reconstruct high-resolution and high-quality images from a group of low-resolution images about the same scene. SR could break through the resolution limit of image acquisition equipment and can achieve data fusion on pixel level.

It is known that visual attention is hierarchical. Attention could be shift between different levels, for example, from large scale to small scale. Resampling depends on the previous segmentation result may obtain more accurate local information with less noise or outliers to rebuild a model for image segmentation.

3 Color leukocyte image segmentation

The framework of color image segmentation is constructed as Fig. 2. The training procedure is illustrated briefly in lower half of the figure. Over-sampling and resampling procedure have been designed in the method, although in some cases their contribution may be neither necessary nor unique.

Color is a key descriptor that guides attention to object location. In order to extract entire leukocyte, we need to locate the effective samples to group training set. In our

method, positive regions are grouped with pixels of nucleus and cytoplasm of leukocyte (or white blood cell, WBC), and negative regions are formed with pixels of mature erythrocyte (red blood cell, RBC) and background. We consider image segmentation is a two-class classification problem and use negative samples to counteract the impacts of mislabeled samples/or noise in positive samples.

Nucleus could be located firstly since it always deep stained. We first use feature-based model over-segment leukocyte image in HSI color space and sort all color blocks according to their average intensity and record their areas. The subregions with lowest intensity and its area over a preset threshold (more than area of a platelet) could be regard as main part of nucleus of WBC, while the subregions with highest intensity are background.

Then, it is easy to get some cytoplasm pixels by dilating nucleus region according to entropy increase rule (using space-based model mentioned above).

After doing above mentioned, we could dilate all the nucleus regions substantially to eliminate the cytoplasm around the nucleus regions roughly. Then remove all the pixels of dilated nucleus and background pixels from the image. The remaining regions in image could be regarded as coarse regions of RBC. We further remove bright pixels by Otsu's method in green component of coarse RBC regions and extract pure pixels of RBC from the color peaks of those remaining pixels. A location result is shown in Fig. 2b (image no. 64).

Sobel derivative operator could be utilized to calculate gradient value of pixels in image. According to formula (10), we can compute the optimal gradient threshold to select the (positive and negative) candidate sampling regions. By sampling/over-sampling from two-class

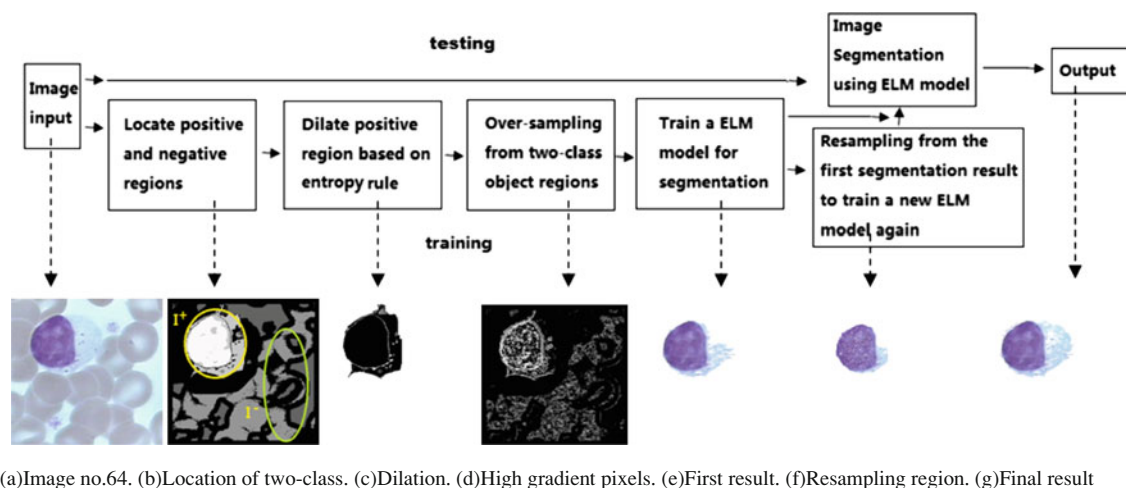


Fig. 2 The framework of the image segmentation and brief illustration for training procedures. Resampling procedure will be skipped when it is unnecessary in experiment

Table 1 Comparison of average performance of three methods

Method	OR (mean)	UR (mean)	ER (mean)	Time-cost (mean) (s)
Watershed-based	0.1661	0.0785	0.2797	9.12
SVM-based	0.0838	0.0518	0.1440	2.61
ELM-based	0.0876	0.0588	0.1567	2.37

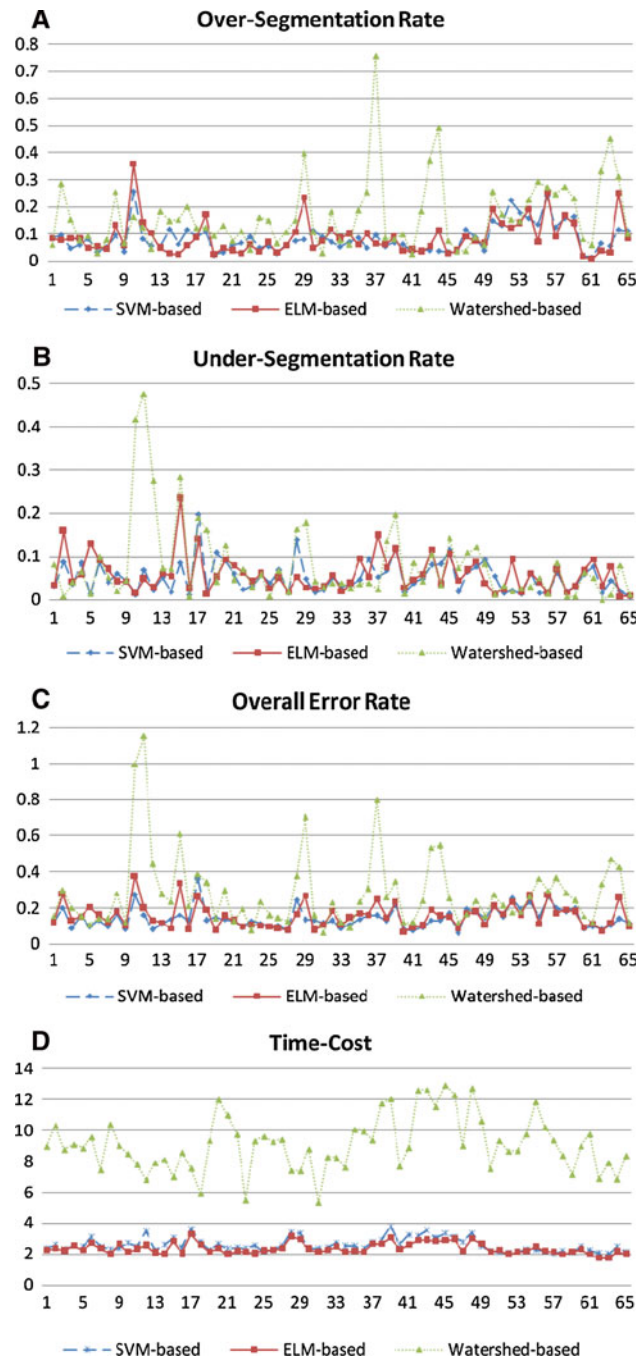


Fig. 3 Comparison of three methods in **a** OR, **b** UR, **c** ER and **d** Time-cost of 65 leukocyte images respectively

regions based on formula (11), we can train an ELM model for image segmentation and then get the first segmentation result.

To obtain more accurate local information without noise or outliers to rebuild a model for image segmentation, we simulate visual system and add a resampling procedure which could group the training set again from the previous segmentation result and produce a new model. According to the sampling rule 2, we need to only sample the high gradient pixels within the objects, so we slightly erode the first segmentation result as candidate regions for resampling. The optimal gradient threshold for selecting high gradient pixels could be calculated again according to formula (10). By resampling from new two-class regions, a new ELM model could be trained and then final segmentation result could be achieved.

4 Experimental results

To demonstrate the validity of the proposed method, 65 blood and bone marrow cell images were tested. These smears stained with Wright-Giemsa method, but acquired from different devices in 2 years with unequal imaging conditions. Since the quality of image may be influenced by many factors, such as uneven staining in the operation, smear thickness, background illumination and maturation of cells, those leukocyte images show complex features, especially in color. It is difficult to segment entire leukocyte using conventional methods. For examples, poor color contrast in cytoplasm of leukocyte may result in weak edge so that edge-based and watershed-based algorithms could not achieve good boundary of leukocyte; color confusion often leads over-segmentation seriously when using thresholding-based approach.

The ELM-based algorithm was programmed with Visual C++ and Matlab7 on a Windows XP-2000 system with a 2.2 GHz dual CPU and 4 GB memory. The contours of the WBC manually drawn by an expert¹ are served as the ground truth. We compared the SVM-based method with our method in same computer. The SVM-

¹ The doctor came from the pathology laboratory of Fourth Military Medical University, Xi'an, China.

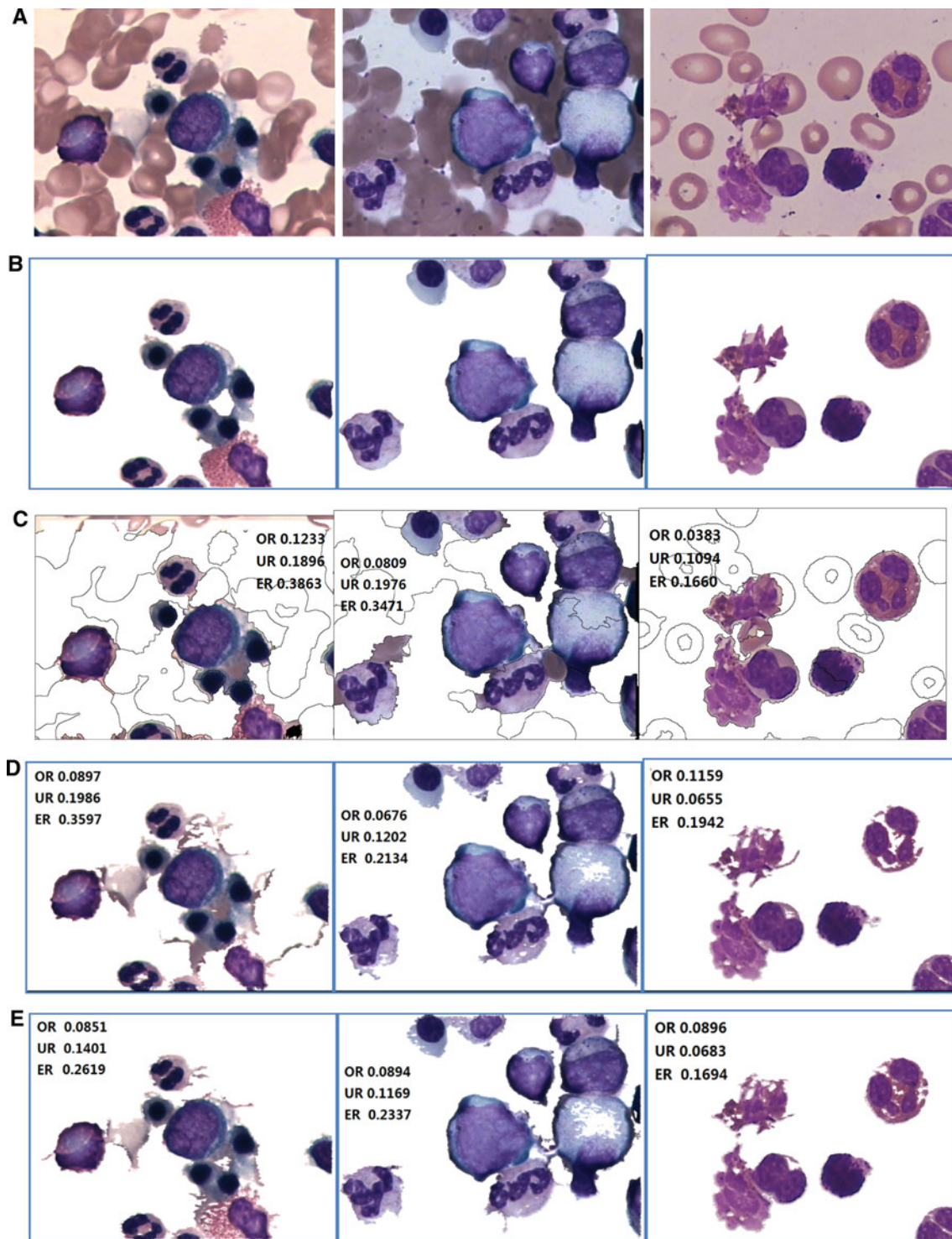


Fig. 4 Some examples in experiment. **a** Original leukocyte images from bone marrow smears. **b** Manual segmentation results as ground truth. **c** Segmentation results based on marker-controlled watershed. **d** Segmentation results based on SVM. **e** Segmentation results based on ELM

based algorithm was implemented with Visual C++ and Libsvm2.8 [3]. SVM used in this experiment is RBF kernel, the parameters are fixed $\gamma = 1$, $C = 100$. Both ELM-based and SVM-based methods share the same procedure of sampling.

Three segmentation error measures will be used to evaluate the performance of segmentation method. Over-segmentation rate (*OR*), under-segmentation rate (*UR*), and overall error rate (*ER*) are often applied to evaluate the ability of a segmentation method in severing the *ROI*

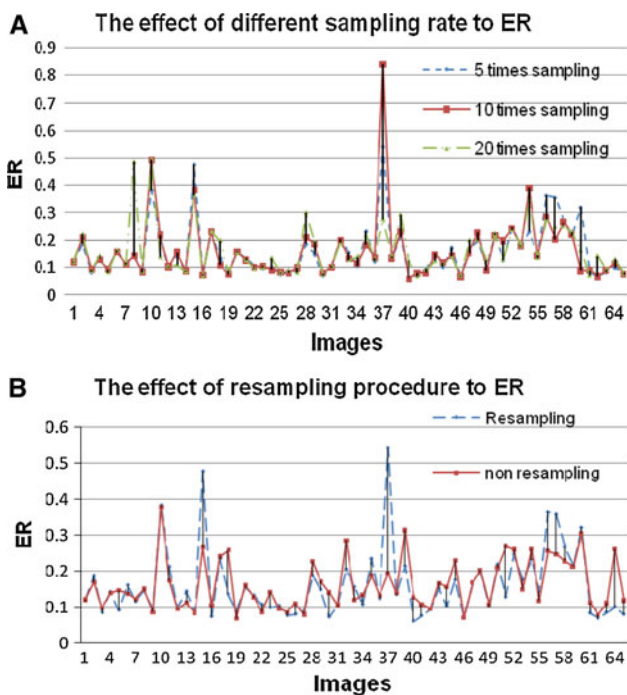


Fig. 5 The effect of over-sampling and resampling to ER. These curves shows high sampling rate, and resampling may reduce the overall error of segmentation effectively in some images. However, they seemly have less influence on most images. The ER performance may deteriorate if noise pixels are sampled into the training set

(region of interest) from an image [11]. Let Q_p be the number of pixels that should be included in the segmentation result but are not, U_p be the number of pixels that should be excluded in the segmentation result but are included, and D_p be the number of pixels that are included in the desired objects generated by manual cutting. Then, OR , UR , and ER can be described as:

$$OR = \frac{Q_p}{U_p + D_p}, \quad UR = \frac{U_p}{U_p + D_p}, \quad ER = \frac{Q_p + U_p}{D_p} \tag{12}$$

Due to ELM is a unified SLFN with randomly generated hidden nodes, the classifier of ELM trained by same training data at different times may be different. So the segmentation results by those classifiers may be slightly different. Although the differences are often a few details near the boundary of object region, it is necessary to maintain the stability of the segmentation. So we can set all the parameters of hidden nodes of ELM with fixed random value.

We design two experiments: One is to compare the marker-controlled watershed, SVM-based and ELM-based methods to demonstrate the effectiveness of our algorithm. The other is to observe the effects of over-sampling and resampling in our method.

4.1 Compare watershed-based, SVM-based and ELM-based methods

The marker-controlled watershed is a popular method to segment cell image. In the first experiment, it could be used as a compared method, and the set of markers that involve object and background are input manually by interaction. We adopt 10 times sampling rate to group training samples for learning-based method. Resampling procedure is skipped in segmentation. Table 1 shows the evaluation of segmentation error and time cost of the three methods. In first three indicators, ELM-based method is slightly lower than SVM-based one. The watershed-based method is lagging in all indicators. The running speed of ELM-based method is quickest, whereas watershed-based method is the slowest. Please note that ELM-based method is hybrid programming using Matlab and VC++, which could be improved further in running speed. Since ELM need not adjust more parameters, it is implemented easily indeed. Figure 3 shows comparison of three methods in OR, UR, ER and time cost of 65 leukocyte images. These curves show that the performances of both SVM-based and ELM-based methods are very close.

Figure 4 provides three instances in the experiment. The first row shows three bone marrow images acquired from different devices with uneven staining and illumination. We can see different types of WBC with various cytoplasm. Manual segmentation is shown in second row as ground truth. The rest three rows show the results of watershed-based, SVM-based, and ELM-based methods, respectively. The watershed-based method shows serious under-segmentation in three images compared to the other methods. ELM-based method produces two better results in Fig. 4 (E-left) and (E-right), while SVM-based method is slightly better in Fig. 4 (D-middle). In our experiment, overall error (ER) could better describe the performance of segmentation. Both SVM-based and ELM-based methods have lower ER, which shows balanced performance of them.

4.2 Observe the effects of over-sampling and resampling

In second experiment, we use ELM-based method to segment images in different sampling rate (5, 10 and 20 times sampling, with resampling procedure). Their ER results are shown in Fig. 5a. Overall from the Fig. 5a, the higher sampling rate reduce the ER in a few images (no. 15, no. 56, no. 57 and no. 60). However, the ER of most images does not decrease in higher sampling rate. It shows that they are in compliance with Nyquist–Shannon sampling theorem. Only in a few images (such as no. 7 and no. 37), the high sampling rates lead to higher ER, which means

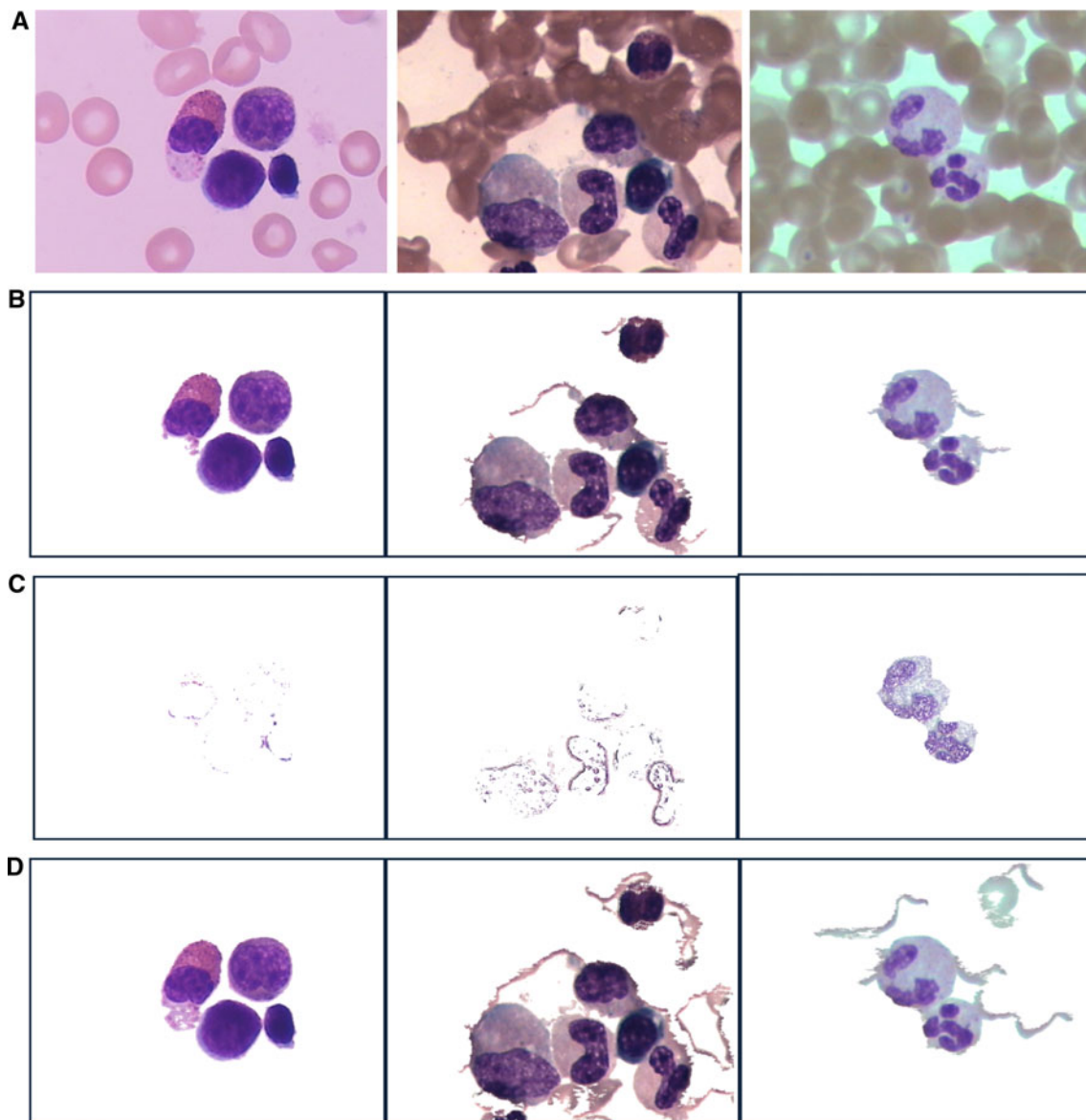


Fig. 6 Segmentation results of image no. 23 (*left*), no. 15 (*middle*) and no. 37 (*right*). By resampling procedure, the segmentation accuracy of no. 23 was improved, while performance degradation

occurs in no. 15 and no. 37 because of color confusion between different objects. **a** Original images. **b** The first segmentation results. **c** The positive resampling regions. **d** The final segmentation results

noise pixels may be sampled into training set results in performance degradation. In this experiment, only HSI color components are used for thresholding to pre-segment image into several subregions with homogeneous color. The color features may be inadequate to a few complex images so that they lead to high under-segmentation error.

Figure 5b shows the difference effects when use or does not use resampling procedure after the first segmentation (in 10 times sampling rate). Resampling reduce the ER value in some images effectively (such as no. 5, no. 23, no. 30, no. 39, no. 40, no. 44, no. 51 and no. 64), which means it is of benefit to improve accuracy of segmentation. Most images maintain their ER value that means resampling has no influence to

those images, and it is seemly unnecessary to those images in this experiment. However, ER becomes higher (performance degradation) in no. 15, no. 37, no. 56 and no. 57, respectively. When the first segmentation results of these images bring noise pixels which may be sampled into final training set by resampling, so that final model trained with impure samples decreases the accuracy of segmentation. We observed those images with higher ER appear more serious color confusion among different objects.

Figure 6 shows one successful example and two failure examples after the process of resampling.

Since segmentation is depended on CLUT model in RGB space in our method, it is hard to overcome the color

confusion when using a CLUT to classify all the pixels of an image.

In order to avoid performance degradation in our method, a preprocessing stage to group pixels into grouped pixels (superpixels) [5, 9, 15] by over-segmentation techniques may be of benefit to segmentation. Based on grouped pixels rather than every pixel, the local and coherent information in superpixel and most of the structure necessary for segmentation at the scale of interest could be preserved. It could reduce probability of under-segmentation in first segmentation and improve the system performance in accuracy and speed. A detail comparison of different over-segmentation techniques used in our framework will be reported in a separate work.

Another idea is to construct many localized models rather than one model for image segmentation. For example, every cell region could be regarded as an adaptive attention window (AAW). Image segmentation will be performed within the AAW. Along the way of this idea, our method could be localized in some AAWs of single cell. This may be more relevant to human visual behavior. We can segment single cell in clustered cells by a localized visual attention-based method. Our future work is to construct a system with multi-level visual attention to get dynamic global and local information to segment natural image. Obviously, few parameters, good generalization performance, and fast training speed of ELM could bring benefits to machine learning-based approach.

5 Conclusions

This paper presents a framework with learning by sampling for leukocyte image segmentation. This idea has been less studied in literature, probably due to the large time needed in training of learning machine. By simulating visual system, thresholding, morphological dilation in local region and ELM algorithm are combined together. Visual attention mechanism guides and limits information processing in efficient way. Although ELM is supervised approach, only a few training samples need be selected and non parameters need be adjusted, so the training of ELM is dramatically fast. Every cell image could be segmented automatically by a special model via ELM training online. Experimental results demonstrate that the new method could extract entire leukocyte from complex scenes, has equivalent performance compared to SVM-based image segmentation, and exceeds the marker-controlled watershed algorithm.

Acknowledgments This research was financially supported by the Ministry of Education, Science Technology (MEST) and National Research Foundation of Korea (NRF) through the Human Resource Training Project for Regional Innovation, the grant Second stage of Brain Korea 21.

References

1. Barlett PL (1998) The sample complexity of pattern classification with neural networks: the size of the weights is more important than the size of the network. *IEEE Trans Inf Theory* 44(2): 525–536
2. Beucher S, Meyer F (1992) *Mathematical morphology in image processing*, vol 12. Marcel Dekker, New York
3. Chang CC, Lin CJ (2001) LIBSVM: a library for support vector machines (Software available at <http://www.csie.ntu.edu.tw/~cjlin/libsvm>)
4. Feng GR, Huang G-B, Lin QP et al (2009) Error minimized extreme learning machine with hidden nodes and incremental learning. *IEEE Trans Neural Netw* 20(8):1352–1357
5. Fergus R, Perona P, Zisserman A (2003) Object class recognition by unsupervised scale-invariant learning. *Proc IEEE Comput Soc Conf Comput Vis Pattern Recognit (CVPR)* 2:264–271
6. Huang G-B, Zhu Q-Y, Siew C-K (2006) Extreme learning machine: theory and applications. *Neurocomputing* 70(6): 489–501
7. Huang G-B, Ding X, Zhou HM (2010) Optimization method based extreme learning machine for classification. *Neurocomputing* 74(12):155–163
8. Jorge OM, Xoana GT, Stephen LM et al (2008) Saccades and microsaccades during visual fixation, exploration, and search: foundations for a common saccadic generator. *J Vis* 8(14):1–18
9. Kadir T, Brady M (2001) Saliency, scale and image description. *Int J Comput Vis* 45(2):83–105
10. Lim YW, Lee SU (1990) On the color image segmentation algorithm based on the thresholding and the fuzzy c-means techniques. *Pattern Recogn* 23(9):935–952
11. Liu J, Leong TY, Chee KB et al (2006) Set-based cascading approaches for magnetic resonance (MR) image segmentation (SCAMIS). In: *Proceedings of AMIA annual symposium*, pp 504–508
12. Matrin R (2009) Microsaccades: Small steps on a long way. *Vis Res* 49(20):2415–2441
13. Otsu N (1978) A threshold selection method from gray level histogram. *IEEE Trans Syst Man Cybernetics* 8(1):62–66
14. Pan C, Fang Y, Yan XG, Zheng X (2006) Robust segmentation for low quality cell images from blood and bone marrow. *Int J Control Autom Syst* 4(5):637–644
15. Ren XF, Malik J (2003) Learning a classification model for segmentation. In: *Proceedings of 9th IEEE international conference on computer vision (ICCV)*, vol 1, pp 10–16
16. Sun YR, Fisher R (2003) Object-based visual attention for computer vision. *Artif Intell* 146(1):77–123
17. Wang ST, Wang M (2006) A new algorithm NDA based on fuzzy cellular neural networks for white blood cell detection. *IEEE Trans Inf Technol Biomed* 10(1):5–10
18. Yeu CWT, Lim MH, Huang G-B et al (2006) A new machine learning paradigm for terrain reconstruction. *IEEE Trans Geosci Remote Sens* 3(3):383–386