Damped Newton based Iterative Non-negative Matrix Factorization for Intelligent Wood Defects Detection

Yafeng Zheng, Qiaorong Zhang
College of Computer and Information Engineering, Henan University of Finance and Economics Zhengzhou, China

Zhao Zhang
School of Information Technology, Nanjing Forestry University, Nanjing, China

***Abstract*—The Non-negative matrix factorization (NMF) can be formulated as a minimization problem with bound constraints. NMF is capable to produce a region- or part-based representations of the wood images. We present an extension to the NMF and discuss the development as well as the use of damped Newton optimization approach for update matrices *W* and *H* called iterative DNNMF with good convergence property for wood defects detection by adding a diagonal correction to the stiffness matrix and employing a Newton direction in the line search until any constraints become active. We also provide algorithms for computing these new factorizations and the supporting theoretical analysis. DNNMF is tested with color wood images based on the statistical features extracted by local binary pattern (LBP) from the feature spaces. Finally, we present experimental results that explore the properties of the proposed method. After many comparative experiments, the test results show DNNMF is effectual and practical with good research values and potential applications.**

***Index Terms*— nonnegative matrix factorization, feature spaces, damped Newton, wood image representation, defects detection, local binary pattern**

I. INTRODUCTION

The Non-negative matrix factorization (NMF) is an un-supervised method whose aim is to find an approximate factorization $V = WH$ to obtain a reduced representation of image data. NMF differs from other methods by its use of non-negativity constraints. However, in fact, non trivial, sometimes the nonnegative factorizations do not always exist as we expected. In the convergence, matrices W and H are initialized with random non-negative values before the iteration starts. In view of the property, various efforts have focused on alternate approaches for initializing or seeding the improved algorithm in order to speed up the convergence of NMF algorithm of Lee and Seung or otherwise influence convergence property to a desired solution [1]. Lin [2] proposed a projected gradient bound-constrained optimization method that is computationally competitive and appears to have better convergence than regular NMF. Zdunek and Cichocki [3] proposed a quasi-Newton approach for the updating rules W and H at the expense of a significant increase in runtime per iteration. The authors of [4] and [5] introduced the concept of approximate and get approximate factorization for NMF by employing the I-divergence between the non-negative matrices and some other approaches. The NMF and some improved NMF approaches have been widely used in data mining fields for which the constraints are relevant, such as image classification [6][7], text mining [8], face recognition [9] and object characterization [10], etc.

The grading of the woods is mainly determined by the defects on wood surfaces and determines the potential uses and values for the Sawmills. Currently, automatic defects detection and the grading of products are one of the key interests and hot issues in the mechanical wood industry. Matti Niskanen [11] and P. Meinlschmidt [12] have adopted the different methods to study wood based defects detection problems.

This paper proposes a new variation to NMF and we take NMF as a framework for wood defects detection, getting help from the damped Newton to reformulate the problem of the NMF and can assure the convergence of nonnegative factorizations by constructing a symmetrical matrix $Q^{(k)}$ to take place of the gradient $\nabla^2 f(h^k)$. The remaining parts of this paper can be organized as follows: In section III, the updating rules for W and H and the detailed convergence analysis will be given. We perform DNNMF on analyzing the wood data and take it for wood defects detection in the section IV by adding the original spatial structure texture features extracted via LBP [13]. Finally, we conclude this paper in Section V.

II. NONNEGATIVE MATRIX FACTORIZATION

For fixed positive image $V_{n\times m}$, NMF's aim is to find a non-negative base $W = [w_{ij}]_{n\times k}$ and matrix $H = [h_{ij}]_{k\times m}$. Lee and Seung [4][15] choose Euclidean distance as the cost function and the NMF factorization is a solution to the following optimization problem:

$$\min_{W,H} f(V, WH) = \frac{1}{2}\sum_{ij}\left(V_{ij} - [WH]_{ij}\right)^2 \tag{1}$$

$$\textit{subject to } W_{ia} \ge 0, H_{aj} \ge 0, \forall i, j, a$$

Here, Eq.1 is a standard optimization problem. The most popular approach to solve Eq.1 is multiplicative

Corresponding author: Yafeng Zheng

doi:10.4304/jsw.5.8.899-906

update algorithm [14]. It is simple to implement and can often yield the good results. For each iteration, the elements of *W* and *H* are multiplied by certain factors. As the zero elements are not updated, all the components of *W* and *H* are strictly positive per iteration. For $1 \le i \le n,\ 1 \le j \le m$ and $1 \le a \le k$, the cost function of NMF of Lee and Seung is non-increasing under the following update rules:

$$W_{i*a} \leftarrow W_{i*a}\left(\frac{(VH^T)_{i*a}}{(WHH^T)_{i*a}+eps}\right) \tag{2}$$

$$H_{a*j} \leftarrow H_{a*j}\left(\frac{(W^TV)_{a*j}}{(W^TWH)_{a*j}+eps}\right) \tag{3}$$

Where, eps is a small positive value to avoid zero in the denominators of the approximations for *W* and *H*. The Euclidean distance is invariant under these updates if and only if *W* and *H* are at a stationary point of the distance. One can normalize the columns of the basis matrix *W* to the unity norm. Then, the column vectors of *W* can be mapped to surface of a hyper-sphere. The detailed proofs of these theorems can be found in [14] and will be extended for damped Newton based case in next sections. Noticing the non-negativity constraints on the matrices *W* and *H* are automatically satisfied by these updating rules if the starting matrices W_0 and H_0 are all non-negative.

III. DAMPED NEWTON APPROACH FOR NONNEGATIVE MATRIX FACTORIZATION

A. Iterative update rules

DNNMF performs on the same cost function as regular NMF, but with the different optimization solutions. For $1 \le i \le n,\ 1 \le j \le m$ and $1 \le a \le k$, the cost function is non-increasing under the following updating rules:

$$H_{a*j} \leftarrow H_{a*j}\left(\frac{\alpha_{(k)}(W^TV)_{a*j}+(1-\alpha_{(k)})(W^TWH)_{a*j}+\beta H_{a*j}}{(((W^TW)_{a*a}+\beta I_{a*a})H_{a*j})+eps}\right) \tag{4}$$

$$W_{i*a} \leftarrow W_{i*a}\left(\frac{\alpha_{(k)}(VH^T)_{i*a}+(1-\alpha_{(k)})(WHH^T)_{i*a}+\beta W_{i*a}}{W_{i*a}(((HH^T)_{a*a}+\beta I_{a*a}))+eps}\right) \tag{5}$$

Where, eps is similarly a small positive value defined to avoid zero(s) in the denominators of the approximations for *W* and *H*. The Euclidean distance is invariant under these updates if and only if *W* and *H* are at a stationary point of the distance.

B. Proofs of convergence

In this paper, we also formally consider algorithms for solving the same optimization problem as [3]: Given a non-negative matrix V_{n*m}, aim to find two non-negative matrices W_{n*r} and H_{r*m} such that: $V_{n*m} \approx W_{n*r}H_{r*m}$ and minimize the cost function of Eq.1.The inequalities such that variables are upper- and lower-bounded are referred to as the bound constraints. So, the problem is a standard bound-constrained optimization problem. We note that

$$\sum_{i=1}^{n}\sum_{j=1}^{m}\left(v_{ij}-(Wh)_{ij}\right)^2 = \|v-Wh\|^2 \tag{6}$$

Where, $\|.\|$ is L2-norm and the normalization procedure can be explained by the defined as that the cost function $f(V;WH) = f(h)$ splits in n independent sub-problems related to each column of the error matrix. Therefore, we consider the partial cost function for a single column of *V*, *W* and *H*, which are denoted by *v*, *w* and *h* respectively:

$$f(h) = \frac{1}{2}\|v-Wh\|^2 = \frac{1}{2}(v-Wh)^T(v-Wh) \tag{7}$$

Let h^k be the current approximation of the minimizer of $f(h)$, then we can rewrite $f(h)$ as

$$f(h) \approx f(h^k) + (h-h^k)^T\nabla f(h^k) + \frac{1}{2}(h-h^k)^TW^TW(h-h^k) \tag{8}$$

Where $\nabla f(h^k) = -W^T(v-Wh^k)$ and W^TW is the corresponding Hessian matrix. So as to receive the steady points, we order $\nabla f(h) = \nabla f(h^k) + \nabla^2 f(h^k)(h-h^k) = 0$. If $\nabla^2 f(h^k)$ is reversible, we can get the following update rules of Newton approach:

$$d^{(k)} = h^{(k+1)} - h^k = \left(-\nabla^2 f(h^k)\right)^{-1}\nabla f(h^k) \tag{9}$$

Where, $d^{(k)}$ stands for the Newton directions. When the initial point is far away from the minimum point, the orientation may not be descending or in bad directions, all of which makes Newton algorithm can not convergent sometimes, but Damped Newton Methods (DNM) can overcome the singularity by adding a diagonal correction to the stiffness matrix. Hence, in this paper, we add one -dimensional search along the directions of Newton [16], whose iterative formula is $h^{(k+1)} = h^k + \alpha_{(k)}d^{(k)}$. Then we can adopt the damped Newton to obtain the update rules for nonnegative factorization. Here, we need meet the following formulations for the fix start points as well as the threshold ε (Here, we always set $\varepsilon = 10^{-6}$)and stop repeating if and only if $\|\nabla f(h^k)\| < \varepsilon$:

$$f(h^k + \alpha_{(k)}d^{(k)}) = \min_{\alpha} f(h^k + \alpha d^{(k)}) \tag{10}$$

Where $\alpha(\alpha \le 1)$ is the corresponding step size after one -dimensional search Here, in order to receive the best step size $\alpha_{(k)}$ for the iterative rules, here we also order $\alpha = 1/2^n, n = 0,1,...,N$ and do $h^{(k+1)} = h^k + \alpha d^{(k)}$ until α satisfy $f(h^{(k+1)}) < f(h^k) - \varepsilon\alpha[\nabla f(h^k)]^T d^{(k)}$ .If so, we will regard the corresponding damping factor α_k as the proper step size for iteration. The iterative procedure of DNNMF is illustrated in Algorithm 1.

Algorithm 1. A damped Newton approach for bound-constrained optimization.

0	Compute the step size α using line search;
1	Initialize $\varepsilon = 10^{-6}$, $\alpha_{(k)} = 1$, $k = 0$, $h^{(k)}$;
2	2.1 If $\|\nabla f(h^k)\|_2 < \varepsilon$, stop;
	2.2 Else, repeat $h^{(k+1)} \leftrightarrow h^k + \alpha_{(k)}d^{(k)}$;
	2.3 If $\alpha_{(k)} = \chi^{n_k}$ and n_k is the first non-negative integer n for which $f(h^{(k+1)}) < f(h^k) - \alpha\left(\nabla f(h^k)\right)^T d^{(k)}$;
	2.4 Else, set $\alpha_{(k)} \leftarrow (1/2)\alpha_{(k)}$;

In addition, how to select the good $d^{(k)}$ for the search direction is very important. We know a good direction $d^{(k)} = \left(\nabla^2 f(h^k)\right)^{-1}\nabla f(h^k)$ needs to satisfy the following two formulations:

$$\left(\nabla f(h^k)\right)^T d^{(k)} \leq -\lambda_{\min} \| \nabla f(h^k) \|^2 \tag{11}$$

$$\| d^{(k)} \| \leq \lambda_{\max} \| \nabla f(h^k) \| \tag{12}$$

Here let λ be the eigenvectors of $\nabla^2 f(h^k)$ and λ meets $0 < \lambda_{\min} \leq \lambda_1 \leq \lambda_2 \leq \ldots \leq \lambda_n \leq \lambda_{\max}$. Eqs.11 and 12 can easily be proved. If now, we just need to make sure $\nabla^2 f(h^k)$ is positive definite, then it is reversible and then we can solve it by $\left(\nabla^2 f(h^k)\right)^{-1} \nabla f(h^k)$. While, for $\forall x \neq 0$, $x^T W^T Wx = \| Wx \|_2^2 \geq 0$, which means $\nabla^2 f(h^k)$ is symmetrical and positive semi-definite. Here, in order to make sure $\nabla^2 f(h^k)$ positive definite, we adopt the similar and effective method as that of [17] to construct a symmetrical matrix $Q^{(k)} = W^T W + \beta I$,where β is a positive number and I is the unity matrix [16].When number β is set properly from zero to tens, we can find an excellent value for it and can guarantee $Q^{(k)}$ symmetric reversible, which can make sure $d^{(k)} = \left(-(W^T W + \beta I)\right)^{-1} \nabla f(h^k)$ is descending and the factorization is global convergent.

Next, the update rules are based on a technique which minimizes $f(h^k)$ by an approximate simpler auxiliary function $Q(h, h^k)$ and satisfy the following formulations $f(h) = Q(h,h)$ and $Q(h, h^T) \geq f(h)$.

$$Q(h,h^k) \approx f(h^k) + (h-h^k)^T \nabla f(h^k) + \frac{1}{2}(h-h^k)^T D(h^k)(h-h^k) \tag{13}$$

Where, $D(x) = diag(x)$ is a diagonal matrix with the vectors x as its diagonal input elements. And here, the formulation $Q(h,h) = f(h)$ is easily verified, so we will just prove $Q(h^T, h) \geq f(h)$ in detail. $D(h^k)$ is a diagonal matrix used to make $(D(h^k) - (W^T W + \beta I))$ semi-definite, which implies $Q(h, h^T) - f(h) \geq 0, \forall h$ .Here, the choice for matrix $D(h^k)$ is similar to that of [14].

$$D(h^k) = diag\left(\frac{\left(W^T W + \beta I\right) h^k}{h^k}\right) \tag{14}$$

Here, we can make sure $D(h^k) - W^T W$ positive semi-definite. If that case, we have:

$$f(h^k) = Q(h^k,h^k) \geq \min_h Q(h,h^k) = Q(h^{(k+1)},h^k) \geq f(h^{(k+1)}) \tag{15}$$

Which can make sure that $f(h^k)$ is non-increasing for h that can be updated by $h^{k+1} = \arg\min G(h,h^k)$. And H can be updated by minimizing $\| V - WH \|^2$ with W fixed. To minimize $f(h^k)$, we update h by $h^{k+1} = \arg\min G(h,h^k)$. And Δh can be resolved by computing $\nabla_h Q(h, h^k)$:

$$\begin{aligned} \Delta h = h^{(k+1)} - h^k &= \left(-D\left(h^k\right)\right)^{-1} \nabla f\left(h^k\right) \\ &= diag\left(\frac{h^k}{\left((W^T W + \beta I)h^k\right)}\right)\left(W^T\left(v - Wh^k\right)\right) \\ &= h^k \frac{\left(W^T v - W^T W h^k\right)}{\left((W^T W + \beta I)h^k\right)} \end{aligned} \tag{16}$$

Besides, we know that $h^{(k+1)} = h^k + \alpha_{(k)} d^{(k)}$, so we can rewrite the components of this equation explicitly as

$$\begin{aligned} h^{(k+1)} &= h^k + \alpha_{(k)}\left(h^k \frac{\left(W^T v - W^T W h^k\right)}{\left((W^T W + \beta I)h^k\right)}\right) \\ &= h^k + \left(h^k \frac{\alpha_{(k)}\left(W^T v - W^T W h^k\right)}{\left((W^T W + \beta I)h^k\right)}\right) \\ &= h^k \left(\frac{\alpha_{(k)}(W^T v) + (1-\alpha_{(k)})(W^T W h^k) + \beta h^k}{\left((W^T W + \beta I)h^k\right)}\right) \end{aligned} \tag{17}$$

Putting together the update rules for all the columns of H yields the desired result for the whole matrix H . By reversing the roles of W and H , in the same way, we can get the corresponding update rules for W and $f(h^k)$ can similarly be shown to be non-increasing under the update rules for W and here can be represented as follows:

$$w^{(k+1)} = w^k \left(\frac{\alpha_{(k)}(vH^T) + (1-\alpha_{(k)})(w^k HH^T) + \beta w^k}{\left(w^k (HH^T + \beta I)\right)}\right) \tag{18}$$

We present our algorithms in the Algorithm 2. DNNMF can be proved convergent per iteration. Firstly, it uses the inverse of the Hessian matrix $\nabla^2 f(h^k)$ to receive the Newton direction by constructing a symmetrical positive definite matrix $Q^{(k)}$. Whenever the rank r of the factor matrices W and H is small, using the inverse Hessian matrix can be advantageous for problems. Secondly, the smallest error ε , size α and the positive number β are made the input parameters, and DNNMF can guarantee monotonic descent of objective function for a sufficiently small ε and a proper step size α .

Algorithm 2. Damped Newton Approaches for NMF.

1 Inputs: r s.t. $1 \leq r \leq \min\{n,m\}$, $V = [x_{i,j}]_{n \times m} = [X_1, X_2, ..., X_m]$,
where, $X_i \in \{R_+^{n*m} / R_+ = [0, +\infty)\}$;
2 Initialize ε , $\alpha_{(k)}$ =1, $k = 0$, $\beta_{(k)}$ =0, $h_{(0)}$, W_{n*r} and H_{r*m} ;
3 Compute the step size α using line search;
  3.1 If $\| \nabla f(h^k) \|_2 < \varepsilon$, stop; Else, repeat
  3.2 Do $h^{(k+1)} \leftrightarrow h^k + \alpha_{(k)} d^{(k)}$;
  3.3 If $f(h^{(k+1)}) < f(h^k) - \alpha [\nabla f(h^k)]^T d^{(k)}$
  set $\alpha \leftarrow \alpha_{(k)}$;
  3.4 Else, set $\alpha_{(k)} \leftarrow (1/2)\alpha_{(k)}$;
4 4.1 If $\left(\nabla^2 f(h^k)\right)^{-1}$ exists, set $\beta \leftarrow \beta_{(k)}$ and
  4.2 do $W^{new} \leftarrow W^{n*r}$; $H^{new} \leftarrow H^{r*m}$;
  4.3 Else, repeat
    4.3.1 Do $\beta_{(k)} \leftarrow i * \beta_{(k)}$, $i = 10 * s$, $s \in R_+$;
    $Q^{(k)} \leftarrow W^T W + \beta_{(k)} I$;
    4.3.2 Till $\nabla^2 f(h^k)$ is positive definite, set
    $\beta \leftarrow \beta_{(k)}$, label;
  4.4 If label is true, loop step 4.2; Else turn to step 4.3.1.
5 $k \leftarrow k+1$;
6 Repeat the steps 4-5 to obtain the updating rules;
7 Outputs: $W \in R_+^{n*r}$, $H \in R_+^{r*m}$, $F \in R_+^{n*m}$;

Where, size α is computed by step 3 and it can be shown that steps 3-4 can decrease the objective function monotonically by the descending directions $d^{(k)}$ .And β is computed by step 4, which can assure $\nabla^2 f(h^k)$ positive definite and the objective convergent. In DNNMF, for the purpose of making sure the gradient descending and proving convergence each time, we need do some indispensable judgments for faster convergence. At the same time, the computational complexity is improved and the processes cost a little more time than that of NMF when facing one image data whose dimension is much larger. However, for the lower dimension images with pixels 30*30, they are almost equally matched. Here, the training set, feature extraction, and detection process are all built on pixels 30*30, under which the runtime gap between DNNMF and NMF methods is very close.

IV. EXPERIMETNS AND ANALYSIS

Wood is an important natural resource and is much in demand. Here, we also present a solution for wood knot defects detection and describe the performances by giving quantitative experiments. The block diagram of SVM [18] based detection system is given in Figure 1. In this paper, the image pretreatments, feature extraction and the defects detection are all built on the matrices, all of which make enough spaces for nonnegative matrix factorization.

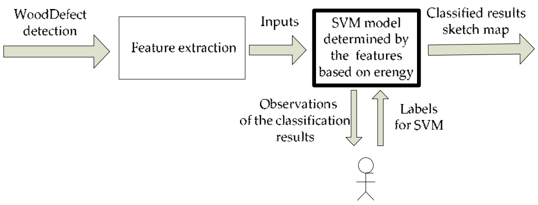

Figure 1. The block diagram of SVM based detection system.

A. *Wood image representation*

Here, we first take some typical wood samples denoted by DEFECT_T from the wood image database [19] provided by VTT Building Technology to examine our algorithm for visualization by comparing with the regular NMF. For NMF, the amount of the base image, k, is also an important factor for the test effects (advantages) between them when facing different values for k. After many cross verifications, we always set k=80 with good generalization ability in the experiments and Figure 2 gives the feature images in the NMF and DNNMF feature subspaces respectively.

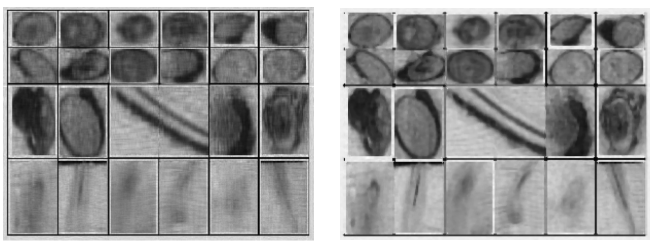

Figure 2. The obtained feature images by NMF(Left panel) and DNNMF (Right panel) for $k = 80$ on the DEFECT_T test set with different sample sizes from pixels 34*32 to 231*223.

From Figure 2, after the NMF and DNNMF processes, the defect regions stand out and wood surfaces are smooth with less disturbed information in the right panel than the left one, and the shapes (including the points and surfaces, etc.) and texture distributions of defects in the right panel are more obvious, all of which makes it possible to detect the defects effectively in the experiments.

In order to intuitively evaluate the high performance of DNNMF for wood image representation, the similarity between the original image and the corresponding feature images is discussed. In this section, Euclidean distances and the Mutual information are used as the measures to evaluate the similarity based on LBP feature histograms. Here, five typical wood samples selected from the data base are used for examples in Figure 3.

As a scaling measure of the link extent of between two signals or images, mutual information can describe the statistical link extent for two random variables. Formally, the mutual information $MI(x, y)$ of two discrete random variables X and Y can be defined as:

$$MI(x, y) = \sum_{x \in X} \sum_{y \in Y} p(x, y) \log \left(\frac{p(x, y)}{p(x)p(y)} \right) \tag{19}$$

with $p(x, y) = \dfrac{c(x, y)}{\sum_{x^T, y^T} c(x^T, y^T)}$, $p(x) = \dfrac{c(x)}{\sum_{x^T} c(x^T)}$, $p(y) = \dfrac{c(y)}{\sum_{y^T} c(y^T)}$

Where, $p(x, y)$ is the joint probability distribution of x and y, and $p(x)$ and $p(y)$ are the marginal probability distributions of x and y respectively. $MI(x, y) >> 0$ shows strong degree of correlation between x and y. $MI(x, y) \approx 0$ represents the link between them is weak. $MI(x, y) << 0$ represents that there is no association relationship existed between x and y.

From Figure 3, we can get the analogous circumstances as Figure 2 that the defect regions stand out and the wood surfaces are smoother with less disturbed information. Besides, the texture distribution and contrast of the feature images structured by DNNMF are more obvious and opportune for distinguishing the defect regions from the backgrounds. Table 1 gives the results of the similarity measures based on the LBP feature histograms.

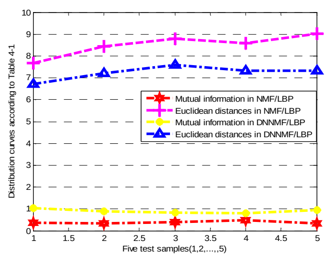

Figure 4. The distribution curves according to Table 1.

From Table 1, we can find the similarities between the original images and the corresponding feature images by DNNMF are closer. Figure 4 also displays the distribution of the curve diagrams according to Table 1 as auxiliary explanation. Considering the Euclidean distances are big and difficult to be drawn in the same figure, we choose the log-form for convenience.

Figure 4 illustrates DNNMF algorithm can describe the characteristics of the wood images better than NMF and can follow the vision point of view of people. Next, LBP feature extraction process is given. LBP is proposed by Ojala [13]. LBP code determined by n sample points is used to represent wood texture features. For fixed pixel centre coordinates (x_d, y_d), LBP is defined as a binary contrast between the center pixel and n surrounding pixels intensity. Texture W is defined as the united distribution of the gray levels of n pixels: $W = w(i_c, i_1, ..., i_n)$, where i_c is corresponding to the gray value of the center pixel of a local neighborhood. i_n (n=1,2,…,P) corresponds to the gray value of n equally spaced pixels on a circle of radius R (R>0) forming a circularly symmetric set.

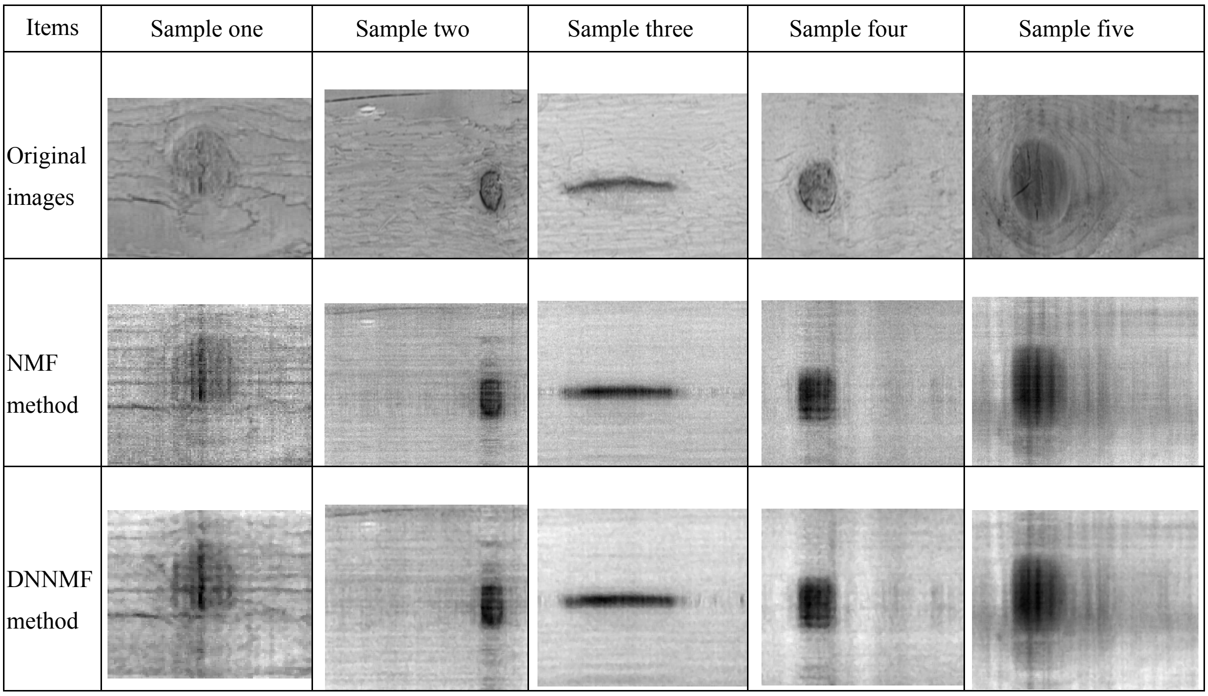

Figure 3. Original wood images and the corresponding feature images obtained in the factorizations.

TABLE I. THE SIMILARITY MEASURES BETWEEN THE ORIGINAL IMAGES AND THE FEATURE IMAGES ON THE LBP FEATURE HISTOGRAMS.

Feature set	**Methods**	**Similarity / Test samples**	**Mutual information**	**Euclidean distances**
LBP feature histograms	NMF/ LBP	Sample one	0.3631	2181
	DNNMF/LBP		1.0517	832
	NMF/ LBP	Sample two	0.3356	4629
	DNNMF/LBP		0.8792	1336
	NMF/ LBP	Sample three	0.3960	6515
	DNNMF/LBP		0.8210	1973
	NMF/ LBP	Sample four	0.4906	5359
	DNNMF/LBP		0.8125	1518
	NMF/ LBP	Sample five	0.3238	8296
	DNNMF/LBP		0.9563	1518

In this paper, we adopt the uniform $LBP_{8,1}^{u2}$ of values (P, R) equal to (8, 1), i.e. around a circle of radius R are eight adjacent pixels and the mapping type is uniform. A simple algorithm for measuring the uniformity of a LBP code is to summarize the absolute value of the difference between the code and the code circularly shifts one bit.

For feature extraction, we first transform color images to gray and divide the images into many small blocks, and then use LBP to extract the texture features described by feature histograms from each block. Finally, we totally obtain 59-dimensional features. The process of the LBP feature extraction is briefly shown in Figure 5.

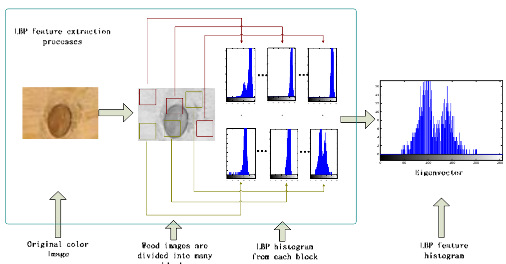

Figure 5. LBP texture feature extraction.

B. Performance standards

Definition 1. For the experiments, the number of detection produced in the system is defined as N_sum. The number of incision on the real defects is defined as N_labeled. The number of incision to the detected and labeled defects in the identification is defined as N_dlabeled and the number of misjudging defects as normal and misjudging sound woods as defects is defined as DN_sum and SD_sum respectively.

Definition 2. The Missed rates indicate the proportion of the missed defects to real labeled defects and can be determined by:

$$M1 = \frac{N\_labeled - N\_dlabeled}{N\_labeled} \times 100\% \tag{20}$$

Definition 3. The Misjudgment rates indicate the proportion of the misjudged defects to the total number of detections.

$$M2 = \frac{DN\_sum + SD\_sum}{N\_sum} \times 100\% \tag{21}$$

Definition 4. Accuracy stands for the detection rate and can be expressed as follows:

$$Accuracy = 1 - M1 - M2 \tag{22}$$

Figure 6 illustrates some but common cases, where, the blue rectangles denote the detections produced in the system. Here, we mainly evaluate the performance of the detection system by *M1* and *M2*.

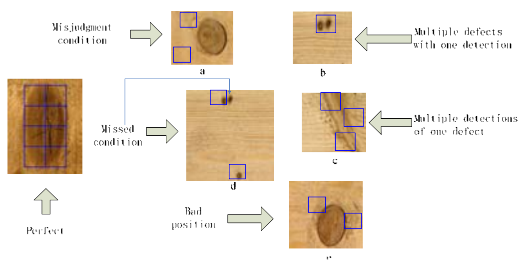

Figure 6. Some examples of detection in the experiments.

C. Wood defects detection

In the experiments, we totally select 1495 samples from the wood image database, in which includes 929 positive samples (labeled by 1) and 566 negative samples (labeled by -1). Figure 7 displays the common seven kinds of knot defects in the actual production. The experiments are based on the direct comparison of NMF with DNNMF for wood defects detection.

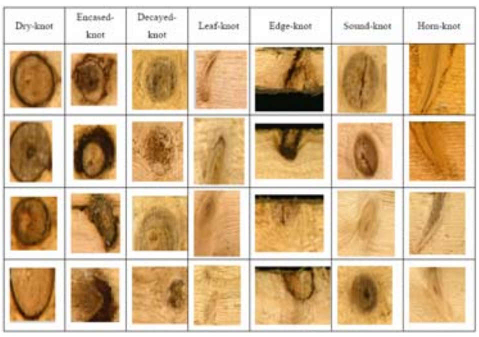

Figure 7. The common seven kinds of knot defects.

Detection system design. The detection system mainly include the following procedures: Building the wood image database for the experiments → Selecting the training set and test sets → Image pretreatments (image format transformation, image segmentation, etc.) → Decompose the wood images by DNNMF → Train the optimal SVM model→ LBP texture feature extraction→ Wood defects detection. Figure 8 has described the main frames of the detection system for convenience.

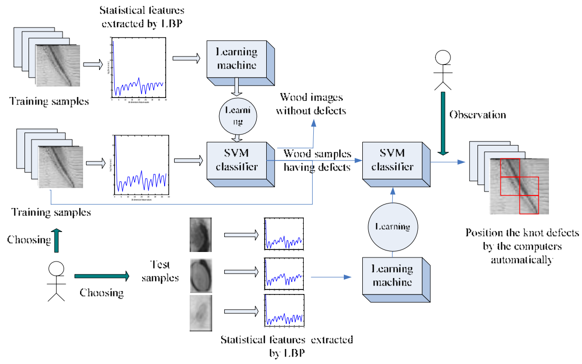

Figure 8. The main frames of the detection system.

Simulation and experiment one. Many wood surfaces are relatively clean and the texture distribution is disciplinarian or some knot defects and the sound woods differ prominently in color depth and texture distribution. However, in the actual production, woods have significant variation both within and between the species. There are no woods that have the same properties in color and texture and surfaces show many varieties of texture characteristics, such as rough, etc. Even for the same species, the defects might greatly vary in shape, size and colors [20]. With large varieties of knot defects and the involvement of human factors cause the current detection methods are unstable and vulnerable to the interferences.

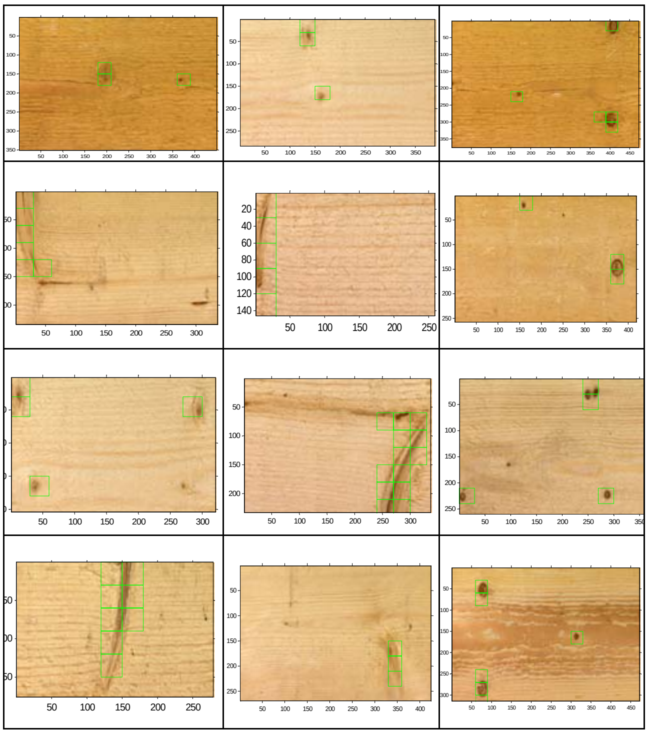

Figure 9. The experimental results of the wood defects detection.

In addition, choosing an appropriate feature set is also very important. After many cross verification, all the feature sets are selected with pixels 30*30 with better generalization. Figure 9 displays the partial test results of DNNMF/ LBP/SVM, in which the boxes are all drawn by computers automatically.Experimental environment: Intel (R) Pentium(R) D CPU 2.80 GHz 2.79 GHz 512.

Simulation and experiment two. In this section, we obey the performance standards defined in section *B* and evaluate the proposed algorithms based on the selected 30 images by comparing with NMF. Table 2 describes the class distribution of defects and the test results observed in the experiments respectively.

From Table 2, we can find the number of leaf knots is smaller than the others. The test results here are evaluated by the *Accuracy* represented by (1-M1- M2), considering one can not effectively judge the types of the knot defects detected, so we will classify it to the type(s) of the current one(s) subjectively

TABLE II.CLASS DISTRIBUTIONS AND THE TEST RESULTS OF THE WOOD DEFECTS DETECTION BASED ON THE DIFFERENT KNOTS.

Results / **Knots**	Dry	Encased	Decayed	Leaf	Edge	Sound	Horn	**Test results (%)**	
								DNNMF/ LBP/SVM	NMF/ LBP/SVM
Dry	45	0	0	0	0	0	0	92.1	90.0
Encased	0	18	0	0	0	0	0	88.5	88.6
Decayed	0	0	20	0	0	0	0	85.6	91.3
Leaf	0	0	0	9	0	0	0	94.8	82.2
Edge	0	0	0	0	21	0	0	86.5	85.9
Sound	0	0	0	0	0	38	0	85.0	88.1
Horn	0	0	0	0	0	0	16	93.5	77.8
Average	$23.85 \approx 24$ (per category)							89.4	86.3

Besides, we find the averaged accuracy of DNNMF/ LBP/SVM exceeds NMF/LBP /SVM, which suggests that the detection system can follow the manual classification even though there exist some unavoidable experimental errors in the processes of wood defects detection.

Figure 10 depicts the schematic diagrams of the test results, from which we can find the gaps of the curves are closer and the curves of DNNMF/LBP/SVM lay below that of NMF/LBP/SVM except for the sound and decayed knots, however, especially predominant on the leaf knots, horn knots and dry knots, all of which demonstrate DNNMF has better convergence and has better detection ability to most kinds of knot defects with LBP and SVM.

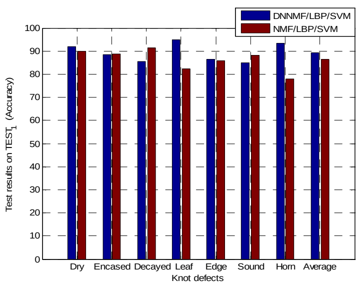

Figure 10.The schematic diagram of the test results.

Table 2 has given prominent to the proposed method and the accuracy can be thought good. Next, we compare the runtime performances of DNNMF/ LBP/SVM with NMF/LBP/SVM based on the selected boards in Table 3.

From Table 3, we can find the total runtime of the proposed method is close to that of NMF/LBP/SVM and the runtime performance can also be thought better. Generally, the runtime performances of the proposed method can be preserved effectively and are able to meet the demands of the practical production better to a certain extent after the wood defects detection algorithm has been improved further.

TABLE III.RUNTIME PERFORMANCE OF THE TWO METHODS.

Methods	Total runtime (s)
NMF/LBP/SVM	48.96
DNNMF/LBP/SVM	52.07

V. CONCLUSIONS AND DISCUSSIONS

The main purpose of this paper is to present a new approach for the regular NMF. The one solving damped Newton sub-problems in Algorithm 2 leads to undoubted convergence than some popular multiplicative update method. Its success is due to our following findings: Sub-problems in Algorithm 2 for DNNMF generally have well-conditioned Hessian matrices (i.e., revised via $Q^{(k)}$) due to the property $r << (m,n)$. Hence, damped Newton method can make sure converge of the objective function. Roughly speaking, damped Newton's method is expensive per iteration to choose a proper step size α and Newton direction d, but has very fast final convergence. With faster convergence or stronger optimization properties, it will be an attractive approach for NMF. Experimental results show the damped Newton's methods are expensive per iteration but have faster final convergence indeed. With the faster convergence and the stronger optimization properties, it is also an attractive approach for NMF.

This paper also applies the DNNMF to wood defects detection and recognition and mainly aims to present a new wood defects detection method, not detection and recognition results. However, the result of the proposed method presented here is good or considerably better.

ACKNOWLEDGMENT

This work was supported in part by Science Technology Project of Henan Province of China under Grant (No. 082102210088).

REFERENCES

[1] Wild, S., Curry, J., Dougherty, A. Motivating Non-Negative Matrix Factorizations. In: Proceedings of the Eighth SIAM Conference on Applied Linear Algebra, July 15-19. SIAM, Williamsburg, VA, 2003.

[2] Lin, C.-J., Projected gradient methods for non-negative matrix factorization. Tech. Rep. Information and Support Services Technical Report ISSTECH-95-013, Department of Computer Science, National Taiwan University, 2005b.

[3] Zdunek, R., Cichocki, A., Non-negative matrix factorization with quasi-newton optimization. In: Proc. Eighth International Conference on Artificial Intelligence and Soft Computing, ICAISC. Zakopane, Poland, June 25-29, 2006.

[4] Lorenzo Finesso, Peter Spreij. Approximate Nonnegative Matrix Factorization via Alternating Minimization. Proceedings of the 16th International Symposium on Mathematical Theory of Networks and Systems, Leuven, July 5-9, 2004.

[5] Michael W. Berry, Murray Browne. Algorithms and Applications for Approximate Nonnegative Matrix Factorization. Computational Statistics & Data Analysis, vol. 52, issue 1, pp.155-173, 2007.

[6] D. Guillamet, B. Schiele, and J. Vitri`a, “Analyzing nonnegative matrix factorization for image classification,” in Proceedings of the 16th International Conference on Pattern Recognition (ICPR ’02),vol.2, pp.116-119, Quebec City, Canada, August, 2002.

[7] D. Guillamet, J. Vitri`a, and B. Schiele, “Introducing a weighted nonnegative matrix factorization for image classification,” Pattern Recognition Letters, vol. 24, no. 14, pp.2447-2454, 2003.

[8] V. P. Pauca, F. Shahnaz, M. W. Berry, and R. J. Plemmons, “Text mining using nonnegative matrix factorizations,” in Proceedings of the 4th SIAM International Conference on Data Mining (SDM ’04), pp.452-456, Lake Buena Vista, Fla, USA, April, 2004.

[9] D. Guillamet and J. Vitri`a, “Nonnegative matrix factorization for face recognition,” in Proceedings of the 5th Catalan Conference on Artificial Intelligence, Castello de la Plana, Spain, October, pp.336-344, 2002.

[10] J. Piper, V. P. Pauca, R. J. Plemmons, and M. Giffin, “Object characterization from spectral data using nonnegative matrix factorization and information theory,” in Proceedings of the AMOS Technical Conference, Maui, Hawaii, USA, September, pp.1-12,2004.

[11] Matti Niskanen, Olli Silv, and Hannu Kauppinen, Machine Vision and Intelligent System Group. Color and textur based wood inspection with nonsupervised clustering, Machine Vision and Applications, pp.275-285, 2003.

[12] P. Meinlschmidt, Wilhelm-Klauditz-Institut (WKI), Fraunhofer-Institute for wood research, Braunsch weig. Thermo-graphic detection of defects in wood and wood-based materials.14th international Symposium of nondestructive testing of wood, Hannover, Germany: May 2nd-4th, 2005.

[13] Ojala T, Pietikäinen M & Mäenpää T. Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. IEEE Transactions on Pattern Analysis and Machine Intelligence,24(7),pp.971-987,2002.

[14] D. D. Lee and H. S. Seung. Algorithms for non-negative matrix factorization. In Advances in Neural Information Processing 13, 2001.

[15] D. D. Lee and H. S. Seung, Learning the parts of objects by non-negative matrix factorization. Nature. 401, pp.788-791, 1999.

[16] D. D. Lee and H. S. Seung, Learning the parts of objects by non-negative matrix factorization. Nature. 401, pp.788-791, 1999.

[17] Jin-de Wang. Stochastic Programmin [M]. Nanjing University Press. Nanjing.1990.

[18] Vapnik,V.N., The Nature of Statistical Learning Theory. New York: Springer-Verlag, 1995.

[19] Sommardahl O, Usenius A. Wood image data, VTT Technology. We are trying to release the data via our WWW pages by the time the paper is published, 1999.

[20] Moring, I. and Silven, O. Automatic Visual Lumber Inspection: Experiments, Finnish Artificial Intelligence Symposium STeP-86, vol.2, pp.290-294, 1986.

Yafeng Zheng was born in Luoyang, China, in May 1979. She received her BS degree in 2002 and Master degree in 2005 in computer application technology from Northwest University, Xian China.

She is a lecture at College of Computer and Information Engineering at Henan University of Finance and Economics, Zhengzhou, Henan, China.

Ms. Zheng’s research interests and publications have been focused on artificial intelligence and network.

Qiaorong Zhang was born in Zhengzhou, China, in Nov 1978. She received her BS degree in 1999 and Master degree in 2002 in computer application technology both from Harbin Engineering University, Harbin China.

She is a lecture at College of Computer and Information Engineering at Henan University of Finance and Economics, Zhengzhou, Henan, China.

Ms. Zhang’s research interests and publications have been focused on image processing, artificial intelligence and biologically inspired computing.

Zhao Zhang was born in Suqian, Jiangsu Province, China, on June 18, 1984. He received his bachelor degree in computer applied science and technology from Nanjing Forestry University in July, 2008. He was recommended as a postgraduate exempted from the postgraduate entrance examinations and becomes a master candidate in Nanjing Forestry University, Nanjing, China.

Currently, he is studying in Department of Information Science and Technology, Nanjing Forestry University, Nanjing. His current research interests include machine learning, intelligent optimization, pattern recognition and the other related application fields.