Defect detection and recognition based on ADABOOT-SVM integrated model

ZhiKai Liang\textsuperscript{1,a}, ShaoZhong Cao\textsuperscript{2,b}, YuKun Tan\textsuperscript{3,c}

\textsuperscript{1}Beijing Institute of Graphic Communication
Beijing No. two, Xinghua street, China
13051379003, 102600

\textsuperscript{2}Beijing Institute of Graphic Communication
Beijing No. two, Xinghua street, China
13520127931, 102600

\textsuperscript{3}Beijing Institute of Graphic Communication
Beijing No. two, Xinghua street, China
18801017073, 102600

\textsuperscript{a}liangzhikai0812@gmail.com, \textsuperscript{b}chaoshaozhong@gmail.com, \textsuperscript{c}tanyukun@gmail.com

ABSTRACT: As the core component of printing machinery, the surface finish and geometric accuracy of printing drum will have an important impact on the quality of printed matter. However, the use of acid ink, alcohol and other chemical raw materials corrode the drum, leading to local collapse or spots. How to effectively identify the types of drum defects has become an important issue. To solve this problem, a defect detection and recognition framework based on adaboot-SVM ensemble learning model is proposed. The framework is composed of two parts: feature extraction and classifier design. The first part is feature extraction from directional gradient histogram (HOG). In the second part, we construct an ensemble of different SVM classifiers to identify defects. The validity of the proposed model is verified by nine different defects. The results show that the integrated model of adaboot SVM is helpful to improve the recognition accuracy of defects.

1. INTRODUCTION
As an important branch of machine vision technology, visual inspection is a hot research direction in the field of product nondestructive testing in China. Data shows that the domestic market in 2015 has reached 350 million US dollars, accounting for 8.3% of the global market, growth rate is 22.2%, ranking first in the world, China has become the world's third largest machine vision market after the United States and Japan. From 2016 to 2020, the growth rate of China's machine vision market is expected to remain above 20%, and will reach a billion dollar market space. Visual sensor has many advantages, such as large amount of information, non-contact with workpiece, high sensitivity and accuracy, strong anti-electromagnetic interference ability, and so on. It is a hot research field. As the core component of printing machinery, the surface finish and geometric accuracy of printing drum will have an important impact on the quality of printed matter. However, the use of acid ink, alcohol and other chemical raw materials corrode the drum, leading to local collapse or spots. How to effectively identify the types of drum defects has become an important issue.
Scholars at home and abroad have studied and proposed a variety of new feature extraction algorithms from the analysis of feature extraction, machine learning, depth learning and other methods, and achieved good recognition results. However, there are many kinds of defects, and some defects are very similar. The selection of effective feature combination usually depends on the experience of experts. The accuracy of recognition is difficult to guarantee. Therefore, it is of great significance to find a data-based self-learning method to improve the performance of defect recognition. With the rapid development of artificial intelligence technology in recent years, machine learning and its application had a hot topic in the field of artificial intelligence. At present, machine learning has achieved good results in some pattern recognition fields [15-19], such as speech recognition, image classification, natural language processing and so on. At the same time, scholars at home and abroad are committed to introducing machine learning into defect recognition [20-24]. Therefore, this paper uses the gradient histogram of Oriented Gradient proposed in reference [25] to extract the features of defects. At the same time, considering that the idea of ensemble learning can be used to construct an effective combined classifier model, it can be used to replace the soft Max classifier used in most depth learning applications. Based on this, this paper proposes a defect identification method combining HOG and ensemble learning. The specific process of the model is shown in Figure 1. First, noise reduction is done in the preprocessing stage. Then, we create the defect features extracted from HOG images. Finally, in the classifier design stage, a multi-SVM linear combination classifier (MSVMLC) is constructed for classification and recognition.

![Figure 1. flow chart of defect recognition based on HOG and ensemble learning](image)

### 2. Histogram of Oriented Gradients

Directional gradient histogram (HOG) is an image descriptor for object detection which is widely used in computer vision and image processing. This method uses the histogram of gradient orientation (HOG) feature to express the detected object, extracts the shape information of the detected object, and forms a rich feature set.

Compared with other descriptors, HOG descriptors have some key advantages. Because it runs on a local element, it is invariant to geometric and photometric transformations except for better capturing local shape information, which only occurs in larger spatial regions. Moreover, the HOG is obtained in a densely sampled image block, and the spatial position relationship between the block and the detection window is implicit in the calculated HOG eigenvector. MATEC Web of Conferences 2 (2012) 01001, 2 corresponding to the volume and 01001 to the number of the article (replacing thus the page number).

a) image normalization

The main purpose of normalized image is to improve the detector's robustness to illumination, because the detector must be insensitive to illumination in different occasions when the target is captured.

b) computing gradient of images using first order differential

1. **Image smoothing**

For gray-scale images, in order to remove noise points, we usually use discrete Gaussian smoothing template to smooth: Gaussian function smoothing gray-scale images at different smoothing scales. Dalal and other experiments show that the next, the best human detection effect (that is, do not do Gaussian smoothing), making the error rate reduced by about one. Double. The possible reason for not doing smoothing is that smoothing is based on edges, which reduces the contrast of edge information and thus reduces the signal information in the image.

2. **Gradient method for image gradient**
The first order differential processing generally has a strong response to the step of the gray scale. First derivative:

\[
\frac{\partial f}{\partial x} = f(x + 1) - f(x)
\]

For the function \( f(x, y) \), the gradient on its coordinate \((x, y)\) is defined by the following two dimensional column vectors:

The module value of this vector is given in the following form:

\[
\nabla f = \begin{bmatrix}
G_x \\
G_y
\end{bmatrix} = \begin{bmatrix}
\frac{\partial f}{\partial x} \\
\frac{\partial f}{\partial y}
\end{bmatrix}
\]

Because the computation cost of modulus value is relatively large, it can be roughly solved according to the following formula:

\[
\nabla f \approx |G_x| + |G_y|
\]

Dalal et al. used many first-order differential templates for gradient approximation, but the results showed that the template \([-1,0,1]\) was the best. Template \([-1,0,1]\) is used as an example to calculate the image gradient and direction. The gradient in horizontal and vertical directions are calculated by gradient template as follows:

\[
G_h(x, y) = f(x + 1, y) - f(x - 1, y) \forall x, y \\
G_v(x, y) = f(x, y + 1) - f(x, y - 1) \forall x, y
\]

Which represents the horizontal and vertical gradient values of the pixel respectively. The gradient value (gradient strength) and gradient direction of the pixel are calculated.

\[
M(x, y) = \sqrt{G_h(x, y)^2 + G_v(x, y)^2} \approx |G_h(x, y)| + |G_v(x, y)| \\
\theta(x, y) = \arctan\left(\frac{G_h(x, y)}{G_v(x, y)}\right)
\]

c) directional weight projection based on gradient magnitude

Generally, there are three kinds of HOG structures: rectangular HOG (R-HOG), circular HOG and center around HOG. Their units are Block (blocks). Dalal's experiment shows that the detection effect of rectangular HOG and circular HOG is basically the same, while the circumferential HOG is relatively poor. Generally, a block is made up of several cells, each of which is made up of pixels. The gradient direction statistics are done independently in each cell so that the gradient direction is the horizontal axis of the histogram. As we mentioned earlier, the gradient direction can be 0 to 180 degrees or 0 to 360 degrees. But the DALAL experiment shows that for human target detection, the direction range of 0 to 180 degrees of neglect can achieve better results. Fruit. The gradient distribution is then averaged into orientation bins, each of which corresponds to a histogram.

d) Normalization of D (HOG) eigenvectors

Normalize the HOG feature vector in block block. The normalization of eigenvectors in block is mainly to make the eigenvector space robust to illumination, shadow and edge changes. The normalization function used in this experiment is L2-norm.

e) get the final eigenvector of HOG.

3. MULTI SVM CLASSIFIER COMBINATION ALGORITHM

SVM is a classification algorithm based on statistical learning theory. It is widely used in many pattern recognition and data mining tasks. SVM classifier not only has strong generalization ability, but also can classify linear non-separable data by mapping data to high-dimensional space through kernel function. Commonly used kernel functions include linear kernel function, Gauss kernel function, polynomial kernel function and so on. At present, the selection of SVM kernel function is still a difficult problem for a given classification task, and multi-SVM combination is an effective method to solve this problem. Therefore, the combined SVM classifier has the following advantages:
(1) according to the above analysis, the combination of SVM classifier can avoid the problem of kernel function selection.

(2) SVM based on different kernel functions has different performance on the same training set. Combining their output results can improve the classification performance.

3.1 Classifier linear combination algorithm

Multi-classifier combination algorithm is a model fusion method, its basic idea is to assume that different classifiers can provide different aspects of the results of information, through the fusion of these information can provide better support for the final decision-making. Reference [27] proposes a linear ensemble method for depth model, which combines the outputs of different DNN models to train a HMM classifier to achieve better results than a single DNN model. In this paper, a linear combination model of multiple classifiers based on posterior probability is constructed on the basis of reference [27]. The model is described as follows:

Consider a schema classification task z, where Z can be assigned to one of the C schema categories, and the object to be classified has C schema categories \{ω_1, ω_2, ω_3, ..., ω_c \}. \( \mathcal{P} = \{P(x_1), P(x_2), ..., P(x_N)\} \) T ∈ \( R^{N \times C} \) True posterior probabilities for N samples. \( p(x_i) \) ∈ \( R^{1 \times C} \) The true posterior probability of a sample is expressed as

\[
p(x_i) = [p(ω_1 | x_i), ..., p(ω_c | x_i)]
\]

Obviously, \( x_i \) is a posteriori probability of belonging to a class is "1" and vice versa "0". Assuming that there are M classifiers, it can be used. \( p(x_i) = [p_k(x_1), ..., p_k(x_N)]^T \) ∈ \( R^{N \times C} \) Represents the posterior probabilities of K classifiers for N samples. Among them, \( p_k(x_i) \) ∈ \( R^{1 \times C} \) Indicates that the posterior probability of the classifier K is output to the sample \( x_i \).

The basic idea of the linear combination strategy of multiple classifiers is that the decision value of the combination model deciding that a sample belongs to a certain pattern category is obtained by linear combination of the output decision values of multiple classifiers, which can be formally described as

\[
P(ω_j | x_i) = α_1p_1(ω_j | x_i) + ... + α_Mp_M(ω_j | x_i) + b_i
\]

Furthermore, the linear combination model of multiple classifiers can be described as

\[
P_{com}(x_i) = \sum_{k=1}^{M} W_k p_k(x_i) + b^T \tag{1}
\]

In the formula, \( W_k \) ∈ \( R^{c \times c} \) represents the weight matrix of the classifier and \( b^T \) ∈ \( R^{c \times 1} \) is the bias vector. Estimating the parameters of Eq. (1) minimizes the difference between the output value and the real value of the combined model, which can be measured by the least square error, i.e.

\[
\min_{(W, b)} C = \frac{1}{2N} \sum_i ||P_{com}(x_i) - p(x_i)^T||^2 \tag{2}
\]

In parameter estimation, regularization term [23] is usually added to the loss function to avoid over fitting. Considering the convenience of solving the model parameters and increasing the L2-regularization term for the loss function, formula (2) is rewritten as

\[
\min_{(W, b)} C = \frac{1}{2N} \sum_i ||P_{com}(x_i) - p(x_i)^T||^2 + \frac{1}{2N} \sum_k γ_k ||W_k||^2 \tag{3}
\]

In the formula, \( γ_k \) is a regularization parameter and can be selected by cross validation. Formula (3) is a multivariate function for extremum problem, which can be solved by least square method. First, the partial derivative of \( W \), \( b^T \) is solved.

\[
\frac{∂C}{∂W} = 0 \text{ and } \frac{∂C}{∂b^T} = 0
\]
\[
\sum_i \left( \sum_{k=1}^M W_k p_k(x_i^T) + b^T - p(x_i^T) \right) p_1(x_i^T) \\
+ \gamma_1 W_1 = 0 \\
\vdots \\
\sum_i \left( \sum_{k=1}^M W_k p_k(x_i^T) + b^T - p(x_i^T) \right) p_M(x_i^T) \\
+ \gamma_M W_M = 0 \\
\sum_i \left( \sum_{k=1}^M W_k p_k(x_i^T) + b^T - p(x_i^T) \right) = 0
\]

The solution system can be solved as follows: \([Wb^T]=BA^{-1}\). Among them, the values of A and B are respectively:

\[
A = \begin{bmatrix}
A_1 & A_2 \\
A_3 & N
\end{bmatrix}
\]

\[
A_1 = \begin{bmatrix}
P_1^T P_1 + \gamma_1 & \cdots & M \\
\vdots & \ddots & \vdots \\
P_M^T P_M + \gamma_M & \cdots & P_M^T P_M + \gamma_M
\end{bmatrix}
\]

\[
A_2 = \begin{bmatrix}
\sum_i P_1^T (x_i), \cdots, \sum_i P_M^T (x_i)
\end{bmatrix}
\]

\[
A_3 = \begin{bmatrix}
\sum_i P_1 (x_i), \cdots, \sum_i P_M (x_i)
\end{bmatrix}
\]

\[
B = \begin{bmatrix}
p^T P_1, \cdots, p^T P_M, \sum_i p^T (x_i)
\end{bmatrix}
\]

In this paper, four different SVM models are selected, which are SVM based on linear kernel (SVM_LK), SVM based on polynomial kernel (SVM_PK), SVM based on Gaussian kernel (SVM_GK) and SVM based on Sigmoid kernel (SVM_SK). From the effect point of view, the combination of multiple SVM classifiers can avoid the problem of SVM kernel function selection to a certain extent. However, the standard SVM can not give the posterior probability output of the sample. In order to use this linear combination algorithm, we need to find a posterior probability output method of SVM.

### 3.2 A posteriori output of SVM

For pattern classification, the standard SVM can get the decision value of the sample belonging to a certain category, but it can not give the corresponding posterior probability. However, there is a certain correlation between the decision value and the posterior probability. The greater the absolute value of the decision value of the sample is, the higher the reliability of the decision value belonging to a certain category will be. Therefore, the posterior probability of the sample can be obtained by the decision value. A posteriori probability of SVM can be obtained by Platt[29]. The essence of this method is to fit a Sigmoid model by training data, which can map the decision value of SVM to a posterior probability. The Sigmoid model can be expressed as

\[
P(y = 1|f) = \frac{1}{1 + \exp(A \ast f + B)}
\]

In the formula, \(f\) is the decision value of SVM output, and the parameters A and B can be obtained through training. It should be noted that this method can only be applied to binary classification problems. For multi-classification problems, it can be classified as multiple binary classification problems, and the corresponding posterior probabilities can be obtained respectively and then normalized.
4. EXPERIMENT AND CONCLUSION ANALYSIS

The validity of this method is verified by classifying 9 types of pictures, including traces, scratches, trachoma, fingerprints, stains, water stains, trailing marks, trailing, droplets. Among them, scratches, trachoma, drag marks and tails are drum defects and need to be replaced. Traces, fingerprints, stains, water stains, water droplets are non-defective, only need to take corresponding treatment measures according to different types. This experiment is completed on the platform of Ubuntu 16.04 + Python 3.6. At the same time, the HOG feature extraction model is built by using the scikit-image library. The LibSVM toolbox not only facilitates the construction of SVM model, but also obtains the posterior probability output of SVM using the method proposed in [20].

Experiment 1: In order to verify the effectiveness of the HOG feature extraction model, several common feature extraction algorithms are selected to compare. LBP (Local Binary Pattern) feature extraction and Haar-like feature extraction are used to select SVM based on linear kernel function and SVM based on Gaussian kernel function respectively to achieve the best recognition results. Then, the feature is extracted by HOG and finally classified by Softmax classifier. The average classification accuracy of each method is shown in Table 1.

| categories                  | 3   | 5   | 7   | 9   |
|-----------------------------|-----|-----|-----|-----|
| LBP+SVM_LK                  | 67.24 | 68.02 | 62.32 | 53.33 |
| Haar_like+SVM_LKK           | 73.32 | 71.03 | 63.53 | 60.56 |
| HOG+Softmax                 | 89.02 | 86.50 | 83.38 | 80.63 |

The results of Table 1 show that the extracted features based on HOG show good performance in classification tasks. And the attenuation rate of accuracy is lower than that of LBP feature extraction and Haar_like feature extraction.

Experiment 2: In order to verify that the multi-SVM combination model can effectively improve the recognition accuracy, this paper extracts the output value of Experiment 1 as the input feature of each comparative classifier (SVM_LK, SVM_PK, SVM_GK, SVM_SK), and combines the four different SVMs according to the algorithm described in Section 2. (MSVMLC), the experimental results are shown in Table 2.

| Classification number      | 3   | 5   | 7   | 9   |
|----------------------------|-----|-----|-----|-----|
| HOG+SVM_LK                 | 92.07 | 91.88 | 87.32 | 85.33 |
| HOG+SVM_PK                 | 96.33 | 91.50 | 87.53 | 85.50 |
| HOG+SVM_SK                 | 95.80 | 91.50 | 87.20 | 84.66 |
| HOG+SVM_GK                 | 96.00 | 90.08 | 87.34 | 82.23 |
| HOG+MSVMLC                 | 98.50 | 93.36 | 90.38 | 89.63 |

From Table 1 and Table 2, we can see that the SVM based on linear kernel function has a certain degree of improvement in classification accuracy compared with the traditional Softmax classifier, and the combined model proposed in this paper can achieve better results than the other four SVM classifiers. Fig. 2 shows the classification accuracy curves of different classification models. It is easy to see that the classification performance of SVM_PK, SVM_PK, SVM_LK and SVM_SK is very similar, and SVM_SK has the best classification performance. At the same time, by comparing the performance curves in Figure 4, we can see that the combination of the four SVM integration model has achieved better classification accuracy.
Finally, in order to further compare the performance of each classification model, a test set is constructed to test. Table 3 is the classification accuracy of the test set for six different classification models. The combined model can also get better performance than SVM alone.

Table 3. Comparison of classification accuracy rate

| Name            | HOG+SVM_LK | HOG+SVM_PK | HOG++SVM_SK |
|-----------------|------------|------------|-------------|
| accuracy        | 86.88      | 84.85      | 84.98       |
| Name            | HOG+SVM_GK | HOG+MSVMLC | HOG+Softmax |
| accuracy        | 85.00      | 87.98      | 85.25       |

5. CONCLUSION

In this paper, a fusion model based on HOG and ensemble learning is proposed, which is applied to classification and recognition of printing cylinder defects, and is verified by identifying 9 kinds of common cylinder defects. The results of Experiment 1 show that compared with the features extracted in literature [3], literature [4] and literature [8], the features extracted in this paper based on HOG model have higher accuracy; according to the results of experiment 2, it can be concluded that multi-SVM combined classifier model can further improve the performance of defect recognition.

REFERENCES

[1] Freund, Yoav, Schapire, Robert E. A decision-theoretic generalization of on-line learning and an application to boosting[C]// European Conference on Computational Learning Theory. Springer, Berlin, Heidelberg, 1995:23-37.
[2] Schapire R E, Freund Y, Bartlett P, et al. Boosting the Margin: A New Explanation for the Effectiveness of Voting Methods[J]. The Annals of Statistics, 1998, 26(5):1651-1686.
[3] Lienhart R, Kuranov A, Pisarevsky V. Empirical Analysis of Detection Cascades of Boosted Classifiers for Rapid Object Detection[J]. Dagm, 2003, 2781:297-304.
[4] Sun Xuechun, Jiang Xiaonan, Fu Yao, etc. Camshaft surface defect detection system based on machine vision [J]. Infrared and laser engineering, 2013, 42 (6): 1647-1653.
[5] Yu Ling, Wu Tiejun. Integrated Learning: A Review of Boosting Algorithms [J]. Pattern Recognition and Artificial Intelligence, 2004, 17 (1): 52-59.
[6] Blumberg B, Downie M, Ivanov Y, et al. Integrated learning for interactive synthetic characters[J]. Acm Transactions on Graphics, 2002, 21(3):417-426.
[7] Jamshed M, Parvin S, Akter S. Significant HOG-Histogram of Oriented Gradient Feature Selection for Human Detection[J]. International Journal of Computer Applications, 2015, 132.
[8] Joachims T. Making large-scale SVM Learning Practical[J]. Technical Reports, 1998, 8(3):499-526.
[9] Burges C J C. A Tutorial on Support Vector Machines for Pattern Recognition[M]. Kluwer Academic Publishers, 1998.
[10] Chang C C, Lin C J. LIBSVM: A library for support vector machines[J]. 2011, 2(3):1-27.
[11] Tomasi C, Manduchi R. Bilateral filtering for gray and color images” ICCV[J]. Proc.ieee Inter.conf.computer Vision, 1998:839.
[12] Zhang B, Allebach J P. Adaptive bilateral filter for sharpness enhancement and noise removal.[J]. IEEE Transactions on Image Processing A Publication of the IEEE Signal Processing Society, 2008, 17(5):664-78.
[13] WANG C, WANG J, ZHANG X. Automatic radar waveform recognition based on time-frequency analysis and convolutional neural network[C]/ Proc. of the IEEE International Conference on Acoustics, Speech and Signal Processing, 2017:2437-2441.
[14] ZHANG M, DIAO M, GUO L. Convolutional neural networks for automatic cognitive radio waveform recognition[J]. IEEE Access, 2017, 5:11074-11082.
[15] LECUN Y, BENGIO Y, HINTON G. Deep learning[J]. Nature, 2015, 521(7553): 436-444.
[16] BENGIO Y, COURVILLE A, VINCENT P. Representation learning: a review and new perspectives[J]. IEEE Trans. on Pattern Analysis and Machine Intelligence, 2013, 35(8):1798-1828.
[17] ABDEL-HAMID O, MOHAMED A R, JIANG H, et al. Convolutional neural networks for speech recognition[J]. IEEE Trans. on Audio Speech and Language Processing, 2014, 22(10):1533-1545.
[18] SAINATH T N, MOHAMED A R, KINGSBURY B, et al. Deep convolutional neural networks for LVCSR[C]/ Proc. of the IEEE International Conference on Acoustics, Speech and Signal Processing, 2013:8614-8618.
[19] HE K, ZHANG X, REN S, et al. Deep residual learning for image recognition[C]/ Proc. of the Computer Vision and Pattern Recognition, 2016:770-778.
[20] VINCENT P, LAROCHELLE H, LAJOIE I, et al. Stacked denoising autoencoders: learning useful representations in a deep network with a local denoising criterion[J]. Journal of Machine Learning Research, 2010, 11(12):3371-3408.
[21] MASCI J, MEIER U, CIREŞAN D, et al. Stacked convolutional auto-encoders for hierarchical feature extraction[C]/ Proc. of the 21st International Conference on Artificial Neural Networks, 2011: 52-59.
[22] DENG L, PLATT J C. Ensemble deep learning for speech recognition[C]/ Proc. of the Annual Conference of the International Speech Communication Association. Singapore, 2014:1915-1919.
[23] NG A Y. Feature selection, L1 vs L2 regularization, and rotational invariance[C]/ Proc. of the 21st international conference on Machine learning. Alberta, Canada, 2004: 78.
[24] PLATT J C. Probabilistic Outputs for support vector machines and comparisons to regularized likelihood methods[J]. Advances in Large Margin Classifiers, 2000, 10(4):61-74.
[25] ZADROZNY B, ELKAN C. Transforming classifier scores into accurate multiclass probability estimates[C]/ Proc. of the 8th ACM SIGKDD international conference on Knowledge discovery and data mining. Alberta, Canada, 2002: 694-699.