Defect Recognition in Radioscopic Image Sequences based on Bag-of-Words

Fan Chang\textsuperscript{1,a*}, Fumin Chen\textsuperscript{1,b}, KaiXuan Tang\textsuperscript{1,c}

\textsuperscript{1}Mechanical Engineering of Xi’an Jiaotong University, Xi’an, China
\textsuperscript{a*}email: changfan1996@stu.xjtu.edu.cn, \textsuperscript{b}email: fmchen@mail.xjtu.edu.cn, \textsuperscript{c}email: think2014@stu.xjtu.edu.cn.

Abstract: Digital X-ray real-time imaging is a new technology in non-destructive detection using image sequences. At present, manual identification of defects is mainly used in practice, which is expertise-dependent and prone to errors. The existing automatic recognition method based on radiographic images mainly recognizes a single defect type for a single radiographic image, which is easy to produce a large number of false detections in non-defect areas, and cannot correctly identify images containing multiple types of defects. A defect recognition method based on Bag-of-Words is proposed, which uses the inter-frame difference method to obtain the moving target area, namely the defect area, and uses the SIFT operator based on the Bag-of-Words model to extract the defect features, and then the trained support vector machine (SVM) performs recognition. Experimental results show that the proposed method can effectively reduce false detections in non-defective areas. The defect recognition method correctly recognizes the defect category with a probability of 85.3%.

1. Introduction
Various defects often occur in the casting process, which seriously affect the quality of the product and pose safety hazards [1]. Common internal defects of castings are shrinkage porosity, shrinkage cavity, bubbles, slag inclusion, and cracks, as shown in Figure1 a) to e) respectively. It is significant for repairing the castings and improving casting quality to identifying the type of defects.

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{figure1.png}
\caption{Defect types in castings}
\end{figure}

In the field of non-destructive testing, radiographic inception has a wide range of applications due to its high sensitivity, high detection efficiency, and the ability of visually display detects [2]. The digital x-ray real-time detection technology can obtain real-time ray image sequences [3]. As shown in Figure 2, the vehicle drives the castings to rotate and move axially, after x-ray irradiation, radioscopic image sequence is captured and transfer to the computer.
At present, many scholars have researched on the automatic defect recognition technology of radiographic images. LiLi Jiang [4] uses a weakly supervised convolutional neural network model to identify the types of casting defects; Mery [5] uses a convolutional neural network (CNN) combined with a sliding window to detect the types of defects in a single casting radiographic image. However, the defect recognition method based on deep learning requires a large number of samples to achieve the satisfied effect, and has high requirements for computer hardware, which is difficult to apply to the detection and recognition of real-time ray image sequences.

A defect recognition method based on Bag-of-Words model is proposed. First, determine changes between the current frame and the previous frame, then do the same on the current frame and the next frame. Perform bitwise AND operation on two binary images obtained above. The boundary of the result represents the defect area. Second, features are extracted from defects, then SVM is used for identifying the defect type. After experimental testing, the proposed method can accurately find the defect area and identify the defect type, and can satisfy the real-time requirements. The technical route of this article is shown in Figure 3.

2. The Method

2.1 Defect Region Extraction

2.1.1 Potential Defect Region Extraction: The inter-frame difference method [6] is often used in the detection of moving targets in video image processing. Its principle is to obtain the contours of moving targets through determining the changes between two adjacent frames. In the radiographic image sequences, the defect area can be regarded as the foreground target, and the defect-free area can be regarded as the background, so we can use the inter-frame difference method to obtain the contours of defects. In order to improve the accuracy of defect detection, firstly, determine changes between current frame and previous frame according to Eq.(1), then determine changes between current frame and next
frame according to Eq.(2), finally, calculate bitwise AND between binary image obtained above according to Eq.(3).

\[ d_{k-1}^k(x,y) = \begin{cases} 1, & f_k(x,y) - f_{k-1}(x,y) \leq T \\ 0, & f_k(x,y) - f_{k-1}(x,y) < T \end{cases} \]  
(1)

\[ d_{k+1}^k(x,y) = \begin{cases} 1, & f_{k+1}(x,y) - f_k(x,y) \leq T \\ 0, & f_{k+1}(x,y) - f_k(x,y) < T \end{cases} \]  
(2)

\[ d_k = d_{k-1}^k \& d_{k+1}^k \]  
(3)

Where, \( f_k(x,y) \), \( f_{k-1}(x,y) \), \( f_{k+1}(x,y) \) represent the current frame, the previous frame and the next frame; \( T \) represent the threshold; \( d_{k-1}^k \) represent the difference image between current frame and previous frame, as shown in Figure 4 a), \( d_{k+1}^k \) represent the difference between current frame and next frame, as shown in Figure 4 b), \( d_k \) represent the bitwise AND result on \( d_{k-1}^k \) and \( d_{k+1}^k \), as shown in Figure 4 c).

![Figure 4 Inter-frame difference result](image)

After obtaining the difference image of the defect, calculate the minimum bounding rectangle of each enclosed area, and merge according to the distance between the centers of the two rectangles. The potential defect area in binary image is shown in Figure 5a), and the potential defect area in radioscopic image is shown in Figure 5b).

![Figure 5 Potential detect area in binary image and radioscopic image](image)

2.1.2 Defect filtering. In the algorithm for extracting potentially defective regions, the threshold is set to be larger in order to reduce missed detections. However, due to the large noise of the radiographic image and low image contrast, reducing missed detections will inevitably increase false detections. Identifying the authenticity of potential defects can effectively reduce false defects. The gray value contrast between the defect area and its neighborhood is high, while the gray value contrast between the false defect area and its neighborhood is low. Taking the 10-pixel wide area around the potential defect area as the neighborhood, the gray value contrast of the potential defect and its neighborhood can be calculated by Eq. (4). where, \( P(x,y) \) represent the gray value of the pixels in the area after expanding the potential defect area by 10 pixels, \( f(x,y) \) represent the gray value of pixels in the potential defect area.

\[ \text{Contrast} = \frac{1}{MN - WH} \left( \sum_{i=1}^{N} \sum_{j=1}^{W} P(x,y) - 2 \sum_{i=1}^{W} \sum_{j=1}^{H} f(x,y) \right) \]  
(4)
2.2 Defect Recognition

2.2.1 Defect Features Extraction. Defect features form a vector to describe defects. The features of similar defects are similar, and the features of different types of defects are obviously different. The purpose of classifying defects can be achieved indirectly by classifying defect features. From the perspective of feature attributes, the defect features include gray-scale features, texture features, and geometric features [7-9]; from the range of features, the defect features are point features, local features, and global features [10]. Gray-scale features (contrast), geometric features (area, circularity, spatial distance) and point features (SIFT operator) are combined to describe defects.

(1) Contrast. The gray difference between the defect area and its neighborhood, namely the contrast.

\[
\text{Con} = \frac{1}{MN-WH} \left( \sum_{i=1}^{W} \sum_{j=1}^{H} P(x,y) - 2 \sum_{i=1}^{W} \sum_{j=1}^{H} f(x,y) \right) - \frac{1}{WH} \sum_{i=1}^{W} \sum_{j=1}^{H} f(x,y)
\]

(5)

(2) Area. The number of defective pixels in the defective area, namely the area.

\[
S = \sum_{i=1}^{W} \sum_{j=1}^{H} f(x,y)
\]

(6)

(3) Circularity. The circularity of the defect indicates how close the defect area is to the circle, namely the circularity.

\[
C = \frac{4\pi S}{L^2}
\]

(7)

(4) Spatial distance. The average distance between defective areas in the radiographic image, namely the spatial distance.

\[
\text{DIST} = \frac{1}{C^r} \sum_{i=1}^{N} \sum_{j=1}^{N-1} \sqrt{(x_i - x_j)^2 + (y_i - y_j)^2}
\]

(8)

(5) SIFT operator. There are large differences in the scale of similar defects in casting defects, and the difference in thin and thick walls leads to differences in the brightness of the radiographic image, these factors all affect the accuracy of classification. The SIFT feature has strong robust to scaling, rotation and illumination change [11]. The calculation steps are as follows:

**STEP1:** Calculate the convolution of the Gaussian function in different scales and the original image to obtain the scale space, then calculate the difference between the adjacent scale spaces to obtain the Difference of Gaussian (DOG), the calculate equation is defined as Eq. (9). Finally obtain the extreme points in the space as candidate feature points.

\[
D(x,y,\sigma) = (G(x,y,k\sigma) - G(x,y,\sigma)) * I(x,y)
\]

\[
= L(x,y,k\sigma) - L(x,y,\sigma)
\]

(9)

**STEP2:** Locate extreme points, use interpolation to obtain the true extreme point positions, and eliminate extreme points with poor robustness

**STEP3:** Determine the main direction of the extreme point, calculate the gradient direction and amplitude of all pixels in the image in the 16×16 neighborhood of the extreme point, and calculate the histogram about the gradient direction. The direction represented by the maximum value of the histogram is the extreme point main direction.

**STEP4:** Generate feature point descriptors, based on the main direction of the extreme points obtained by STEP3, divide the 16×16 neighborhood around the feature points into sixteen 4×4 regions, then calculate the gradient magnitude and the pixel gradient in each 4×4 region direction, and divide the gradient direction into 8 segments, then calculate gradient histogram. Finally combine 16 histograms to obtain a 128-dimensional feature vector, which is the SIFT feature descriptor.
The bag-of-words model [12] uses a set of independent keywords (visual words) to describe a set of images, and an image can be described by the frequency of these visual words. Using the bag-of-words model can reduce the dimension of the feature vector and enhance the description ability of the feature vector. The principle of the bag-of-words model is shown in Figure 6. The steps are as follows:

**STEP1:** Extract the SIFT feature description operator of each image in the training set and merge it into a feature set.

**STEP2:** Generating distance centers as visual words by K-Means clustering.

**STEP3:** Calculating the frequency of each image's feature description operator belonging to every visual words, then compose the frequency into a vector, and normalize them.

**STEP4:** Combining contrast, area, circularity, spatial distance with the vector obtained from STEP3 into a new feature vector.

**STEP5:** Establish SVM model.

**STEP6:** Train the SVM model using train set, and evaluate the model using the test set.

2.2.2 Defect Features Classification. Support vector machines construct linearly separable problems by introducing a kernel function, which greatly optimizes the calculation method of nonlinear classification problems and reduces the influence of sample dimension on algorithm efficiency. Compared with neural networks and clustering algorithms, support vector machines show strong generalization ability and robustness in a small set of samples, so this paper uses support vector machines as classifiers.

3. Result and Analysis

264 samples are used for training, including 113 shrinkage porosity, 83 shrinkage cavity, 28 bubbles, 30 slag inclusions, and 10 cracks. 80 samples are used for testing, including 25 shrinkage porosity and 25 shrinkage cavity, 10 bubbles, 10 slag inclusions, 5 cracks. The sample distribution histogram is shown in Figure 7 below.

![Sample distribution histogram](image)

Eq. (10) is used to calculate the classification accuracy of each defect type in test sample, where represents the correctly classified quantity in type , and represents the total quantity in type , where represent the correctly classified quantity in type and represent the total quantity in type . The result is shown in Table 1 below:
accuracy = \frac{TP_i}{Sample_i} \quad (10)

| Type              | accuracy |
|-------------------|----------|
| shrinkage porosity| 88%      |
| shrinkage cavity  | 92%      |
| bubbles           | 70%      |
| slag inclusion    | 100%     |
| cracks            | 100%     |

It takes 30 minutes to manually detect a workpiece, with high false-detection and false-recognition. The proposed algorithm takes an average of 10 minutes to detect a workpiece, and the recognition accuracy is 85.3%.

4. Conclusion
Digital X-ray real-time detection is one of important Non-Destructive Testing (NDT) methods for casting defect detection. Manual interpretation of defects is mainly used in practice, which is highly subjective and prone to make false-detection and false-classification. A defect recognition method based on Bag-of-Words model is proposed. Compared with the manual recognition and other automatic recognition method based on single frame, the proposed method can effectively reduce false-detection and have a higher accuracy rate.

Acknowledgements
Very honored to participate in this meeting. Thanks very much Professor Chen for his guidance on this paper.

References
[1] Couper M J, Neeson A E, Griffiths J R. Casting defects and the fatigue behaviour of an aluminium casting alloy[J]. Fatigue & Fracture of Engineering Materials & Structures, 1990, 13(3): 213-227.
[2] Hull J B, John V. Non-destructive testing[M]. Macmillan International Higher Education, 2015.
[3] Shao J, Du D, Chang B, et al. Automatic weld defect detection based on potential defect tracking in real-time radiographic image sequence[J]. NDT & E International, 2012, 46: 14-21.
[4] Jiang L, Wang Y, Tang Z, et al. Casting defect detection in X-ray images using convolutional neural networks and attention-guided data augmentation[J]. Measurement, 2020: 108736. IET Intelligent Transport Systems, vol. 12, Oct. 2018, pp. 939-946, doi: 10.1049/iet-its.2018.5127.
[5] Mery D. Aluminum Casting Inspection Using Deep Learning: A Method Based on Convolutional Neural Networks[J]. Journal of Nondestructive Evaluation, 2020, 39(1): 12.
[6] Gao F, Lu Y. Moving Target Detection Using Inter-Frame Difference Methods Combined with Texture Features and Lab Color Space[C]. 2019 International Conference on Artificial Intelligence and Advanced Manufacturing (AIAM). 2019.
[7] Haralick R M, Shanmugam K, Dinstein I H. Textural features for image classification[J]. IEEE Transactions on systems, man, and cybernetics, 1973 (6): 610-621.
[8] Wang L, Pavlidis T. Direct gray-scale extraction of features for character recognition[J]. IEEE transactions on pattern analysis and machine intelligence, 1993, 15(10): 1053-1067.
[9] Junyan L, Qingju T, Yang W, et al. Defects’ geometric feature recognition based on infrared image edge detection[J]. Infrared Physics & Technology, 2014, 67: 387-390.
[10] Mikolajczyk K, Leibe B, Schiele B. Local features for object class recognition[C]//Tenth IEEE International Conference on Computer Vision (ICCV'05) Volume 1. IEEE, 2005, 2: 1792-1799.

[11] Lowe D G. Object recognition from local scale-invariant features[C]//Proceedings of the seventh IEEE international conference on computer vision. Ieee, 1999, 2: 1150-1157.

[12] Zhang Y, Jin R, Zhou Z H. Understanding bag-of-words model: a statistical framework[J]. International Journal of Machine Learning and Cybernetics, 2010, 1(1-4): 43-52.